Getting Down to FACTS

Academic Gatekeeping of English Learner-Classified Students: The Case of California

Ilana Umansky
University of Oregon

Havisha Khurana
University of Oregon

May 2026

Stanford | SCALE Initiative
Accelerator for Learning

Academic Gatekeeping of English Learner-Classified Students: The Case of California

Ilana Umansky and Havisha Khurana

University of Oregon

Acknowledgements: We want to thank the Learning Policy Institute (LPI) for partnering with us, and in particular Heather Price, Dion Burns, Stacy Loewe, and Linda Darling-Hammond. We are grateful to the California Department of Education for the opportunity to work with statewide data via a Data Sharing Agreement with LPI. We are grateful as well to the leadership of Getting Down to Facts III, especially Susanna Loeb and the other researchers working on studies to support multilingual learners: Lucrecia Santibañez, Alfredo Artiles, Francesca Lopez, and Heather Price. Many thanks too, for the collaboration of CORE-PACE including Alix Gallagher, Belen Chavez, Jeannie Myung, and Lupita Cortez Alcalá.

Introduction and Background

English learner (EL) classification is a legally-required status for multilingual public-school students in K-12 schools who meet federal requirements for language support services. By law, the EL category is designed to be temporary. As they develop proficiency in the English language and no longer need or benefit from supplemental supports, students are exited from the category. The process of exiting EL status is called reclassification and, in California, is governed by a combination of federal, state, and local rules. Because students' status and services shift once exited, reclassification is considered a potentially consequential status change, and a significant body of research has explored the rules, contexts, and effects related to reclassification. This study adds to this body of work, addressing research questions of targeted importance to California.

Specifically, current California law requires that students clear an academic English language arts threshold to be eligible for reclassification. Called the "basic skills" criterion (California Education Code § 313(f), 2025), this requirement is designed to ensure that students who exit EL status are prepared to succeed without supplemental services in academic settings. However, the basic skills requirement also means that students who are fully English proficient may be held in EL status for reasons unrelated to their English proficiency (Novicoff et al., 2025). This study looks specifically at this group of students – those who pass the state's English proficiency criterion – and asks how being held in EL status rather than being reclassified influences their academic and behavioral outcomes.

Drawing on a stratified random sample of 76 California school districts of diverse geographical contexts and EL concentrations, we employ a fuzzy frontier regression discontinuity design to ask: For students who meet California's English proficiency reclassification requirement but are on the margin of the basic skills requirement, how does being held in EL status impact subsequent year test scores, academic proficiency, attendance, and discipline? Further, we explore whether the effects of holding students in EL-status differ for students in districts with more strenuous versus more lenient basic skills requirements. Our results indicate that, on average, there are no statistically significant subsequent-year effects of being held in EL status for this population of students. Among districts with more strenuous academic criteria, however, there is a positive impact of being held in EL status on next-year English language arts (ELA) performance, suggesting a possible initial shock of reclassification and/or an instructional disconnect that leaves recently reclassified students struggling in general education. This report describes those results and addresses implications for California reclassification policy and EL education more broadly.

Policy Context

In California, reclassification policy is the result of federal, state, and local rules. Federal law requires that states use standardized statewide procedures for entry and exit into the EL category; and that, at a minimum, those procedures include an annual assessment of each EL-classified student's English language proficiency (ELP) using an assessment that measures English speaking, listening, reading and writing skills (Every Student Succeeds Act [ESSA], 2015). Federal law leaves up to individual states the selection of ELP assessment, the setting of ELP thresholds on the assessment, and any additional rules related to reclassification.

Most U.S. states use only their ELP exam for to determine reclassification eligibility (33 states), with some requiring a single overall ELP score as the sole reclassification criterion (24 states), while others require that students meet thresholds on the specific proficiency domains (reading, writing, speaking, and listening) within that assessment (9 states; Morales & Lepper, 2024). Other states, such as Florida, and Texas do not limit reclassification criteria to the state ELP assessment but instead require additional criteria such as academic criteria (test scores, proficiency levels, and/or grades) and/or teacher, parent, or administrator approval. While there is suggestive evidence that having these additional, non-ELP related criteria might pose a barrier to students' educational progress (for example, states with non-ELP criteria have considerably larger proportions of students considered long-term ELs; Morales & Lepper, 2024), there is little direct evidence of the impact of having academic or other non-ELP criteria.

California is alone in requiring four separate reclassification criteria areas (Faulkner-Bond, 2026). Specifically, California state statute requires (1) that a student receives a minimum summative proficiency level of four ("Proficient") on the state's ELP assessment, the English Language Proficiency Assessments for California (ELPAC)[1]; (2) teacher evaluation including, but not limited to, an evaluation

[1] This criterion differs in some cases for students with "the most significant cognitive" disabilities (CDE, no date).

of a student's curricular "mastery", an evaluation that is locally determined but that can involve a tool developed by the state called the Observation Protocol for Teachers of English Learners (OPTEL); (3) parental opinion and consultation as per a locally-determined process; and (4) a locally-determined measure of a student's basic skills relative to English proficient students (California Department of Education [CDE], 2025). These criteria are additive, not substitutive, meaning that a student must meet all four criteria to be eligible for reclassification. Also noteworthy, three of the four criteria (criteria two through four) are locally determined by individual school districts, with significant variation across the state (Hill et al., 2021).

This study is focused specifically on the fourth criterion, the basic skills criterion. The CDE describes this criterion as: "Comparison of the performance of the pupil in basic skills against an empirically established range of performance in basic skills based upon the performance of English proficient pupils of the same age, which demonstrates whether the pupil is sufficiently proficient in English to participate effectively in a curriculum designed for pupils of the same age whose native language is English" (CDE, 2025). The use of an academic criterion, like this one, for reclassification is controversial. The arguments in favor of such a criterion, which were leaned on heavily in California's decision to have such a criterion are two-fold. First, having an academic criterion puts pressure on education systems (e.g., schools, districts) to ensure that EL services support academic growth. Related to the notion of 'teaching to the test', the idea is that systems are less likely to overlook the academic needs of EL-classified students if those skills are tested and required for a high-stakes accountability metric (Jennings & Bearak, 2014). The second argument in favor of an academic criterion is that it helps ensure that reclassified students will be ready and successful in reclassified status without EL services. Stakeholders want to be sure that students are not being reclassified prematurely only to struggle or worse, fail, in their general education setting (Linquanti, 2001).

While compelling, the arguments against such a criterion are equally persuasive. Closely connected to the first argument in favor of an academic criterion, an argument against is rooted in a recognition that EL-classified students may not always be provided equitable access to rigorous grade-level academic content (Umansky et al., 2024). Indeed, federal law allows for English development and academic content to be provided sequentially. In other words, systems may delay full access to content while providing English development supports (U.S. Department of Justice & U.S. Department of Education, 2015). As such, students subject to an academic reclassification criterion may be personally penalized (i.e. fail to pass requirements for exit) not due to any fault of their own but instead because of omissions in the content that they were taught. A true catch-22, students may then be trapped in EL status because the status itself results in an inability to exit.

The second main argument against an academic criterion is that it sets up a barrier to educational access that is unique to EL students. Non-EL students have a full distribution of academic performance, some score high, others score low, but they are not removed from general education content due to a low academic score. EL-classified students subject to an academic criterion, however, do face such a barrier. They must score above a specific academic threshold (in the case of California, their peers' average performance level) in order to be in general education services. Finally, and perhaps most importantly, critics argue that "English learner" classification by its very definition, is about being in a process of developing English. In this sense, exiting EL status should relate specifically to no longer developing required English skills, and not relate to any requirements that are not direct measures of English proficiency (Linquanti, 2001).

These questions of how and where to set reclassification criteria are important because while reclassification at first glance appears to be an administrative phenomenon there is increasing evidence that reclassified status exposes students to different educational environments, services, and opportunities than EL status, and these differences can shape students' short- and long-term outcomes. California is a particularly important state to examine reclassification effects not only because of its complex multi-measure approach to reclassification, but also because of the magnitude of the state's EL and multilingual student population. California has the largest number of EL-classified students of any state in the country. Indeed, in California over half of students under age five are growing up in multilingual homes, accentuating the long-term importance of understanding reclassification effects in the state (Williams & Zabala, 2023). With the following research questions, our aim is to help the state think through policy effects of current law, provide data and evidence for any future policy revisions, and inform other states in the U.S. as they also consider reclassification policy. Specifically, we ask:

1) What are basic skills and ELP reclassification criteria in a set of randomly sampled California districts?
2) Among our sample of districts, how many students meet reclassification criteria for basic skills alone, ELP alone, both basic skills and ELP, and how many were subsequently reclassified? How does this vary by subgroup?
3) Accounting for district specific criteria, what is the impact being held in EL status on academic (test scores and proficiency rates) and behavioral (attendance, suspension) outcomes among otherwise eligible students who pass the ELP criteria and are on the margin of the basic skills criterion? How does this impact vary by grade band?
4) What, if any, are the differences in the effects of being held in EL status at the margin of the basic skills criterion for students in districts with lenient versus strenuous basic skills criteria?
5) Among districts with similar basic skills reclassification criteria, is there meaningful heterogeneity by district in the effects of being held in EL status?

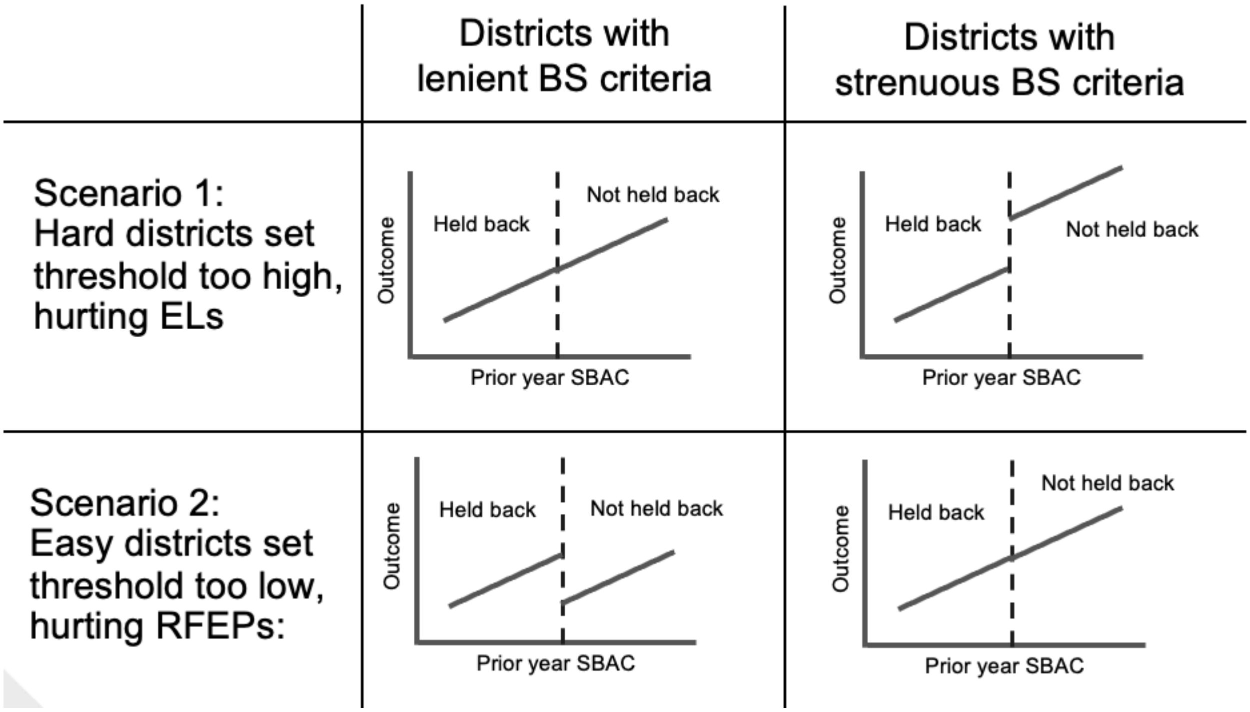
Two Conceptual Frameworks for Understanding Reclassification Effects

Studies that examine how reclassification, versus remaining in EL status, impact students are often used to assess the appropriateness of the level or levels at which reclassification thresholds are set (Betts et al., 2020; Robinson, 2011). Embedded within this framework (which we call the *conventional framework*) is the assumption that there is a correct level at which to set reclassification threshold(s); set too low, the argument goes, and students will enter general education before they are ready and flounder; set too high, by contrast, and students will remain in EL services after they are necessary and stagnate. Set just right, students will smoothly transition from EL to reclassified status. Of note, an assumption underlying the conventional framework is that the educational context of students is held constant or, perhaps, immaterial. Stakeholders can find the “right” (Betts, et al., p.2) reclassification threshold(s) so that students are reclassified at the “right” time and that this is done or can be done independent of any evaluation of or harmonization with services or supports for students.

While typically applied to ELP thresholds, the same conventional framework can be applied to a setting where there are reclassification criteria measuring multiple constructs, such as in California. The inclusion of an academic criterion, in essence, raises the bar for reclassification by adding an additional requirement for student passage. If the academic criterion is set high, the increased difficulty of reclassification is clear: students who meet non-academic criteria may fail to meet the academic criterion and, as such, fail to be eligible for reclassification. But even if set low, the inclusion of an additional criterion generates administrative and logistical complexity that can create barriers to reclassification (Bartlett et al., 2024; Estrada & Wang, 2013). Thus, in a setting like California, one can apply this conventional framework toward the assessment of both the level and presence of an academic reclassification criterion.

Because there is local control over reclassification criteria in California, we can generate a useful illustrative schema (see Figure 1) using this conventional framework. In this schema, we can imagine districts that set their basic skills thresholds high, and others that set them low. Since there is, in the conventional framework, a correct place to set the basic skills threshold, it follows that either (or both) the threshold in the high-threshold districts is too high (hurting students near the threshold who are held in EL status, as shown in Scenario 1) or the threshold in the low-threshold districts is too low (hurting students near the threshold who are reclassified as shown in Scenario 2). Thus, if this framework holds, we expect to find one of the scenarios shown in Figure 1. In this study, we can begin to test the utility of this conventional framework by seeing if our data fits one of these scenarios.

Figure 1: Conceptual Schema Applying Conventional Framework to Understanding Reclassification Effects in a Setting with Variation in Criteria Difficulty



Note. BS = Basic skills. EL = English learner. RFEP = Reclassified fluent English proficient. SBAC = Smarter Balance Assessment Consortium (ELA test).

While the conventional framework is useful from a policy evaluation lens, sociolinguists remind policy researchers that language fluency and proficiency are social constructs that are context- and stakeholder-determined. As such, there is no one 'right' level at which reclassification thresholds should be set (Flores & Rosa, 2015; Valdés, 2005). More recently, then, researchers and practitioners posit instead that reclassification thresholds cannot be considered on their own, but instead are part of a larger interconnected system that includes the types of learning environments that students are in when classified as EL and when reclassified out of EL status. As authors, we agree with this later position, recognizing that an educational setting could successfully lower their reclassification thresholds without impacting student outcomes if they were to ensure that recently reclassified students had ample linguistic and academic supports in their general education environment. Likewise, a setting could raise its reclassification criteria without harming students if it ensured that students, while EL-classified, enjoyed all the academic rigor and curricular opportunities offered to non-ELs (Cimpian et al., 2017).

Given this *alternative framework*, it is less straightforward how to interpret reclassification effects as indication that a given threshold is too high, too low, or just right. Applied to Figure 1, data might not fit neatly into one of these two scenarios. Instead, we might find that being held in EL status (or conversely, being reclassified) had different effects in different settings even when settings had the same or similar reclassification criteria. We might also find patterns not observed in Figure 1 such as positive effects of being held in EL status only in districts with more rigorous criteria, or negative effects of being held in EL status only in districts with easier criteria.

In the alternative framework, the effects of being held in EL status or being reclassified out of EL status should be interpreted *in tandem* with thinking about the educational settings of students before and after reclassification (Cimpian et al., 2017). Specifically, in settings where being held in EL status results in negative educational outcomes, educational settings for these students should be adapted to ensure rigorous and expansive access. By contrast, if being held in EL status benefits students near the reclassification threshold, this suggests that multilingual students are not having their needs met in general education settings. These settings should be examined for accessibility, relevance, and equity for these students alongside any evaluation of reclassification levels.

Existing Research on Academic Reclassification Criteria Effects

Prior literature has explored the impact of being held in EL status versus being reclassified on a range of student outcomes. Synthesized, this body of work finds that reclassification can be consequential in multiple ways, but that the presence and magnitude of impacts differ across settings, populations, timing, and outcomes (Garrett et al., 2022; Johnson, 2019; Ma & Winters, 2024; Onda & Seyler, 2020). A recent meta-analysis finds that across studies there are no consistent, statistically significant effects of being held in status versus being reclassified on academic outcomes or graduation but that there is evidence that reclassification may result in a short-term negative shock on academic outcomes that then dissipates and reverses (becoming positive) over time. This negative shock has been theorized to be the result of changing instructional environments and the removal of specialized supports for EL-classified students once reclassified (Itoh & Umansky, no date).

While most states (33) have relatively simple reclassification policies, relying only on ELP-related criteria (Morales & Lepper, 2024), prior research has, at times, accounted for the fact that reclassification effects may differ based on which criterion is the most challenging for any given student (Robinson-Cimpian & Thompson, 2016). These studies use frontier regression discontinuity designs and can examine, as do Lee and Soland (2022) and Robinson (2011), whether reclassification effects differ based on whether the lowest score of a given student was an English proficiency score versus and academic score. In the above two cases, the authors found that reclassification effects did not vary according to whether students' lowest score was an academic versus an English proficiency score.

To our knowledge, only two studies have focused on reclassification effects related to an academic criterion specifically (Robinson-Cimpian & Thompson, 2016; Xu et al., 2025). Both use a methodology like the one we use here, including only students in their samples who pass English proficiency-related criteria but whose academic ELA scores position them close to the academic reclassification threshold. It is important to qualify that these studies, like ours, are not directly evaluating the impact of having an academic criterion. What they do, instead, is evaluate how reclassification versus remaining EL impacts outcomes for a group of students who, in the absence of an academic criterion, would have been reclassified. Robinson-Cimpian and Thompson (2016) used data from a California school district (pre-ELPAC and SBAC) and identified that reclassification effects at the academic threshold varied based on the level at which the English proficiency criterion was set. Taking advantage of the rescaling of the state's ELP test that effectively increased the level of ELP required for reclassification eligibility, they found negative reclassification effects on test scores and graduation when the ELP threshold was lower, effects that dissipated when the English proficiency bar was raised. This finding speaks to the complex interplay of reclassification criteria and in particular to the importance of where the ELP reclassification threshold is set in relation to students' instructional needs.

The second study, a statewide study in Texas, also found complex and mixed effects of reclassification versus remaining EL for English proficient students at the academic margin.

Reclassification in 8th grade resulted in an increase in the likelihood of taking advanced courses in high school but had a negative effect on course passage rates and graduation. Through a combination of theory and mechanism checks, these authors suggest that the group of students that benefited from reclassification was largely different from the group that was negatively impacted and that the negative impacts were uniquely experienced by students who experienced more extensive changes in instructional settings upon reclassification (Xu et al., 2025).

What do these studies tell us about the merits or dangers of an academic criterion? They confirm that reclassification is consequential, with indication that at least in some settings EL status may constrict course access. But they also point to the fact that EL-related instructional supports for academic success in school (as measured by test scores, course passage, and graduation) can be important for multilingual students, even those who meet ELP and academic reclassification criteria. Perhaps what stands out most about these studies is that they do not lead to clear conclusions about whether reclassification should require an academic criterion and instead suggest that what matters more than a student's language classification is whether they are in an educational setting that responds to their needs and offers them full educational opportunity.

The present study, like the two studies described above, uses a regression discontinuity design to examine, among students who have met state English proficiency standards for reclassification, how being held in EL status based on English language arts (ELA) achievement impacts a set of outcomes. We do so for students at different grade levels, and unlike the prior two studies, we look at both academic and behavioral outcomes. Because California's academic criterion is locally determined, we are also able to stress test the conventional framework for understanding reclassification effects described above, looking both at how effects of being held in EL status differ based on relative basic skills criterion difficulty level, and the degree of variation in effects among districts with the same or similar basic skills criteria. We next describe our data and methods.

Data and Method

Data

In this study, we draw on student level longitudinal data from the California Department of Education (CDE) for the years 2021-22 through 2023-24. We accessed this data via a Data Use Agreement (DUA) in place with the Learning Policy Institute. CDE data include critical variables for our analysis including language status and reclassification timing, SBAC and ELPAC scores, and student information such as race, grade, and school and district of attendance.

While ideally, we would have been able to evaluate the impact of English proficient students being held in EL status over multiple years and across the full state, California's use of locally-determined reclassification criteria limited us to examining *next-year* reclassification effects among a *sample* of districts. This is because in order to estimate the impact of reclassification using a regression discontinuity design, we need to know what exact basic skills reclassification criteria were in place in each locale. California does not maintain a centralized collection of district criteria, nor how they change over time. As such, we needed to collect district-specific criteria manually, limiting the number of districts that we could include. Further, we needed to limit our analysis to the most recent years of data when we could relatively safely assume that reclassification criteria were consistent.

District Sampling

To capture the wide diversity of educational contexts in California, we developed a stratified random sample of districts, stratifying on district locale and multilingual student population. Specifically, we used National Center for Education Statistics (NCES) and CDE public files from the 2021-22 school-year to make a district-level dataset with the following information: district locale (city, town, suburb, rural) (from NCES), and enrollment, number of ever EL-classified[2] students and number of students eligible for free or reduced price lunch (FRPL) (from CDE). Of California's 940 districts, we removed one with missing locale information as well as small districts, operationalized as enrollment

[2] Ever EL is a category of students that includes all students who were at some point in their K-12 experience classified as EL, whether or not they retained that classification. It is calculated by summing current EL students and reclassified students (Thompson et al., 2023).

below the 25th percentile of the state enrollment distribution (equating to districts serving fewer than 384 students). This resulted in the removal of an additional 235 districts.

We then defined twelve strata: the four locale types and within each locale type three categories of EL concentration. The EL concentration categories were based on locale-specific terciles of proportion of school enrollment that was ever EL.[3] Next, from each stratum, we drew 10 random districts (for a total of 120 districts) to include in the potential analytical sample. We also drew an additional 10 districts in each stratum for use should any of our initial sampled districts need to be replaced. Both authors searched each district's website to locate reclassification criteria (often found within the district's EL master plan) between July and September 2025. We made a key assumption that the reclassification criteria found on the website had been in place long enough to accurately represent criteria in the most recent years for which we had student-level data, i.e. the 2021-22 and 2022-23 school years. Note that although every district (excluding those removed, as detailed above) had a non-zero chance of being selected in the sample, the probability varied by locale due to differing district counts by locale. Specifically, the selection probabilities were 1 in 47 for cities, 1 in 44 for rural areas, 1 in 48 for towns, and 1 in 95 for suburbs. Consequently, suburban districts were underrepresented in the sample relative to their actual proportion in the state.

As described above, CDE data includes SBAC test scores, but state statute allows districts to choose which assessment or assessments they use for the basic skills criterion. Because we needed test scores on the basic skills criterion assessment to run our regression discontinuity models, we had to limit our sample to districts that used the SBAC in their basic skills criterion. As a result, we categorized districts into one of four groups (see Table 1) based on our search results. The first category was districts for which we could find reclassification criteria, and which used the SBAC. These districts (N=68) remained in our sample. The second category was districts for which we could find criteria but that did not use the SBAC (N=21). These districts we discarded and replaced with an alternative district in the same strata. The third category was districts for which we could not find reclassification criteria. These districts (N=79) we contacted via email. If we received a response with reclassification criteria within two weeks, we proceeded to evaluate that district's eligibility for inclusion. If we did not receive

[3] Cut points for these terciles were as follows: City: 0.27, 0.43; Town: 0.13, 0.45; Suburb: 0.18, 0.36; Rural: 0.11, 0.29.

a response, we replaced the district with a different within-strata district. The final category was districts for which we could find criteria and which used SBAC, but which were not sufficiently clear in their criteria for us to be able to operationalize their basic skills criteria (N=29). An example of this is a district that described their basic skills criterion as “CAASPP Language Arts Required Score: No lower than 15% below the mean scale score of the English-only students OR- Meets or Exceeds Standard.” In these cases, we emailed a district administrator for clarification on the basic skills criterion. If we received a response within two weeks, we included the district in our sample. If we did not receive a response within that time frame, we replaced the district with another within-strata district. In scenarios 2-4, we only made one replacement. If the replacement district was not eligible, we did not replace a second time. In the end, we kept 76 districts in our sample.

Table 1: Process for Sampling and Replacing Sampled Districts

Condition	Found Master Plan/ Reclassification Criteria?	District Uses SBAC for basic skills criteria?	SBAC Criteria Clear?	Decision
1	Yes	Yes	Yes	Entered analytic sample
2	Yes	No		Replace district with another random district from the strata – replace only once. Email district administrator to ask for EL Master Plan or reclassification criteria.
3	No			If receive response within 2 weeks, follow one of conditions 1 or 2. If no response, replace district in the strata.
4	Yes	Yes	No	Email district administrator for clarification. If receive response within 2 weeks, follow one of conditions 1 or 2. If no response, replace once in the strata.

Note. As detailed in text, the 12 strata were defined by district locale and within-locale terciles of ever EL concentration.

Reclassification Criteria Coding

We then turned to coding reclassification criteria for our sample of districts. The second author used image and text-scraping tools[4] to put together a district-grade level dataset with the original text, as published by the district, related to each of California's four reclassification criteria. Both authors then independently coded each district's ELP and basic skills reclassification criteria. While we collected data for California's reclassification criteria 2 (teacher evaluations) and 3 (parent consultation), we did not code this data because CDE data do not include variables that would allow us to assess individual students' eligibility on either of those criteria. Our codebook is presented in Table A1 in the appendix

[4] We found that most district reclassification criteria were presented at tables or pointers. Therefore, we used `magick`, `EBImage`, and `tidyverse` packages in R to create editable versions of the tables in Excel by identifying table boundaries and extracting cell information. Code is available at https://github.com/havishak/table_img_to_df.git

and includes data elements such as “SBAC Only” which indicates that the district used SBAC ELA performance as the only assessment to determine eligibility on the basic skills criterion and “SBAC Thresh Level” which documented the SBAC proficiency level set as the basic skills criterion threshold for that district in that grade. Any disagreements were reviewed to reach consensus, but there was extensive agreement between the two authors.[5] Once fully coded, we then merged our district-grade reclassification criteria data into our student level data.

Dataset Merging

We kept only student-year observations in the districts in our sample and only observations where the student was classified as an EL in the prior year. We limited observations to the 2022-23 and 2023-24 academic years and to students with valid prior-year SBAC and ELPAC scores. Because SBAC is only administered in grades 3-8 and 11, our sample was limited to students who had valid past-year scores and were in grades 4-9 and 12.[6] The result is a student-year panel dataset where each student is observed up to two times. Of note, we structured each student-year row such that the row includes both current year test scores and prior year test scores. As a result, we can tell by looking at a given student-year observation both what that student’s current year SBAC and ELPAC scores were and what their prior year SBAC and ELPAC scores were.

Table 2 provides descriptive statistics of our sample (column 2). It also shows, in column 1, descriptive statistics for students across the state over two academic years. For comparability with our district sample, the values in the state sample column are similarly comprised of two cohorts of students who were, in the prior year, classified as EL with valid SBAC and ELPAC scores. In total, our district sample includes 112,077 student-year observations, between 55,000 and 57,000 students in each of the academic years, representing just under 11% of the state. Comparing columns 1 and 2 we see that across almost all measures, our random sample is very similar to the full state, suggesting that our results likely represent the state well. Our sample differs from the state in two notable ways: first, it

[5] Initial agreement between authors was 99% as relates to district ELP criteria and 85% as relates to basic skills criteria. After discussion, the authors reached consensus on all discrepancies.

[6] A small number of observations came from students with valid past year scores but who were repeating the same grade in the current year, i.e., repeaters. We retained these observations in the analytical sample, which therefore includes a small number of observations from grades 3 and 11.

slightly over samples Asian students and under samples Latine students, and second, it consists of a lower proportion of students attending rural districts and a higher proportion attending districts located in towns (see Table 2).

Table 2: Student Characteristics across State, Sampled Districts, and the Regression Discontinuity Analytic Sample

	(1) State Sample	(2) Sampled Districts	Regression Discontinuity Analytic Sample*		
			(3) Total	(4) Below Reclassification Threshold	(5) Above Reclassification Threshold
N	1,032,548	112,077	14,455	6,270	8,185
Entered Current-Year as Reclassified					
No	791,144 (76.6%)	87,656 (78.2%)	3,071 (21.2%)	2,145 (34.2%)	926 (11.3%)
Yes	241,404 (23.4%)	24,421 (21.8%)	11,384 (78.8%)	4,125 (65.8%)	7,259 (88.7%)
Current-Year Grade level**					
3	341 (0.0%)	13 (<0.0%)	<10	<10	<10
4	192,545 (18.6%)	21,330 (19.0%)	1,354 (9.4%)	395 (6.3%)	959 (11.7%)
5	190,036 (18.4%)	21,379 (19.1%)	2,662 (18.4%)	914 (14.6%)	1,748 (21.4%)
6	171,772 (16.6%)	19,236 (17.2%)	3,022 (20.9%)	1,292 (20.6%)	1,730 (21.1%)
7	144,864 (14.0%)	16,523 (14.7%)	2,249 (15.6%)	1,077 (17.2%)	1,172 (14.3%)
8	130,938 (12.7%)	14,740 (13.2%)	2,766 (19.1%)	1,339 (21.4%)	1,427 (17.4%)
9	116,848 (11.3%)	9,252 (8.3%)	1,570 (10.9%)	849 (13.5%)	721 (8.8%)
11	1,377 (0.1%)	101 (0.1%)	<10	<10	<10
12	83,827 (8.1%)	9,503 (8.5%)	824 (5.7%)	400 (6.4%)	424 (5.2%)
Academic Year					
2022-2023	525,407 (50.9%)	57,120 (51.0%)	7,117 (49.2%)	2,969 (47.4%)	4,148 (50.7%)
2023-2024	507,141 (49.1%)	54,957 (49.0%)	7,338 (50.8%)	3,301 (52.6%)	4,037 (49.3%)
Gender					
Female	475,339 (46.0%)	51,690 (46.1%)	6,645 (46.0%)	2,770 (44.2%)	3,875 (47.3%)
Male	557,001 (53.9%)	60,377 (53.9%)	7,809 (54.0%)	3,499 (55.8%)	4,310 (52.7%)
Low Socio-Economic Status Flag					
No	117,169 (11.3%)	15,615 (13.9%)	2,661 (18.4%)	1,005 (16.0%)	1,656 (20.2%)
Yes	915,379 (88.7%)	96,462 (86.1%)	11,794 (81.6%)	5,265 (84.0%)	6,529 (79.8%)
Race/Ethnicity					
African American	4,547 (0.4%)	513 (0.5%)	70 (0.5%)	24 (0.4%)	46 (0.6%)
Asian	89,448 (8.7%)	15,448 (13.8%)	2,682 (18.6%)	977 (15.6%)	1,705 (20.8%)
Filipino	10,116 (1.0%)	1,180 (1.1%)	151 (1.0%)	61 (1.0%)	90 (1.1%)

	(1) State Sample	(2) Sampled Districts	Regression Discontinuity Analytic Sample*		
Hispanic	871,506 (84.4%)	88,996 (79.4%)	10,630 (73.5%)	4,836 (77.1%)	5,794 (70.8%)
Other	14,236 (1.4%)	1,908 (1.7%)	269 (1.9%)	114 (1.8%)	155 (1.9%)
White	42,695 (4.1%)	4,032 (3.6%)	653 (4.5%)	258 (4.1%)	395 (4.8%)
Language					
Arabic	13,958 (1.4%)	1,764 (1.6%)	280 (1.9%)	117 (1.9%)	163 (2.0%)
Chinese	22,497 (2.2%)	3,010 (2.7%)	570 (3.9%)	183 (2.9%)	387 (4.7%)
Other	98,769 (9.6%)	14,376 (12.8%)	2,217 (15.3%)	848 (13.5%)	1,369 (16.7%)
Russian	7,498 (0.7%)	592 (0.5%)	127 (0.9%)	49 (0.8%)	78 (1.0%)
Spanish	871,547 (84.4%)	88,833 (79.3%)	10,609 (73.4%)	4,832 (77.1%)	5,777 (70.6%)
Vietnamese	18,279 (1.8%)	3,502 (3.1%)	652 (4.5%)	241 (3.8%)	411 (5.0%)
Special Education					
No	827,075 (80.1%)	90,620 (80.9%)	13,537 (93.6%)	5,785 (92.3%)	7,752 (94.7%)
Yes	205,469 (19.9%)	21,456 (19.1%)	918 (6.4%)	485 (7.7%)	433 (5.3%)
Years in EL					
<=2	20,824 (2.0%)	2,372 (2.1%)	147 (1.0%)	51 (0.8%)	96 (1.2%)
3--5	279,178 (27.0%)	31,714 (28.3%)	3,305 (22.9%)	1,133 (18.1%)	2,172 (26.5%)
5+	732,546 (70.9%)	77,991 (69.6%)	11,003 (76.1%)	5,086 (81.1%)	5,917 (72.3%)
Newcomer Status					
No	954,505 (92.4%)	104,236 (93.0%)	14,074 (97.4%)	6,143 (98.0%)	7,931 (96.9%)
Yes	78,043 (7.6%)	7,841 (7.0%)	381 (2.6%)	127 (2.0%)	254 (3.1%)
Past-Year ELA Performance Level					
1	660,491 (64.0%)	68,821 (61.4%)	2,354 (16.3%)	2,350 (37.5%)	<10
2	246,191 (23.8%)	27,823 (24.8%)	8,399 (58.1%)	3,919 (62.5%)	4,480 (54.7%)
3	103,196 (10.0%)	12,401 (11.1%)	3,620 (25.0%)	<10	3,619 (44.2%)
4	22,670 (2.2%)	3,032 (2.7%)	82 (0.6%)	<10	82 (1.0%)
Past-Year Math Performance Level					
1	713,927 (69.3%)	75,358 (67.4%)	6,212 (43.0%)	3,303 (52.7%)	2,909 (35.6%)
2	209,330 (20.3%)	23,430 (20.9%)	4,966 (34.4%)	1,995 (31.9%)	2,971 (36.3%)
3	78,075 (7.6%)	9,183 (8.2%)	2,328 (16.1%)	739 (11.8%)	1,589 (19.4%)
4	26,004 (2.5%)	3,618 (3.2%)	916 (6.3%)	217 (3.5%)	699 (8.5%)

	(1) State Sample	(2) Sampled Districts	Regression Discontinuity Analytic Sample*		
Missing	3,202 (0.3%)	265 (0.2%)	18 (0.1%)	<10	<10
Student-Level Outcomes					
Std. ELA Score	-0.651 (0.805)	0.071 (0.691)	-0.017 (0.617)	-0.193 (0.600)	0.109 (0.598)
Std. Math Score	-0.584 (0.805)	0.048 (0.781)	-0.044 (0.718)	-0.195 (0.691)	0.064 (0.717)
Prop. ELA Proficiency Met or Exceeded	0.163 (0.369)	0.459 (0.498)	0.402 (0.490)	0.290 (0.454)	0.482 (0.500)
Prop. Math Proficiency Met or Exceeded	0.117 (0.322)	0.308 (0.462)	0.249 (0.432)	0.177 (0.382)	0.300 (0.458)
Prop. Chronically absent	0.475 (0.499)	0.379 (0.485)	0.384 (0.486)	0.414 (0.493)	0.361 (0.480)
Prop. Any Suspension	0.056 (0.230)	0.040 (0.196)	0.039 (0.195)	0.049 (0.215)	0.032 (0.177)
District Characteristics					
Locale					
City	495,751 (48.2%)	53,736 (47.9%)	7,328 (50.7%)	3,185 (50.8%)	4,143 (50.6%)
Suburb	412,834 (40.1%)	42,210 (37.7%)	5,092 (35.2%)	2,190 (34.9%)	2,902 (35.5%)
Town	76,842 (7.5%)	13,899 (12.4%)	1,769 (12.2%)	741 (11.8%)	1,028 (12.6%)
Rural	43,274 (4.2%)	2,232 (2.0%)	266 (1.8%)	154 (2.5%)	112 (1.4%)
%Ever-EL	41.050 (15.031)	36.744 (14.720)	37.121 (15.018)	36.749 (15.080)	37.406 (14.965)
%Low-SES	71.919 (17.700)	65.031 (22.786)	65.895 (22.817)	66.086 (22.419)	65.748 (23.118)
District-Size	72,803.152 (156,597.175)	21,868.770 (18,106.894)	20,786.145 (17,204.992)	20,400.124 (16,542.500)	21,081.851 (17,691.022)

Note. EL = English learner. ELA = English language arts. SES = socioeconomic status. Prop. = proportion. The N size for the state shown in row 1, column 1, of this table includes all students in the 2022-23 or 2023-24 academic years who were, in the prior year, classified as EL, and for whom we have valid prior-year ELPAC and SBAC scores. The N row for our sample shown in row 1, column 2, is the same but limited to student observations in our random sample of 76 districts. Ever-EL refers to all current and former (reclassified) EL students. Low-SES refers to students eligible for free- or reduced-price lunch. Table reports frequencies (percentage) for categorical variables and mean (standard deviation) for continuous variables.

* We define the regression discontinuity (RD) analytic sample as students within the sampled districts who met district ELP Criteria and who fell within 0.75 standard deviations (SD) of the mean of the distribution of our standardized and centered SBAC Score.

**We restricted our sample to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past year.

Key Variables

Running Variable*.* The RD design relies on the assumption that, near the cut point, students who score below the threshold are similar, on average, in every way relevant to either treatment assignment or outcomes, to students who score above the threshold. A key element of any RD study is that treatment eligibility is determined by a given individual's position above or below a known cut point on a continuous measure. This measure, in our case the SBAC ELA assessment, is the running variable. Because districts set their own thresholds on the measure, we centered each student's SBAC score relative to their district-specific reclassification cut-score. In addition to centering these SBAC scores to district-set thresholds, we also standardized scores within year and grade.

Treatment. In a sharp RD design, treatment is defined as being above or below the cutoff of the running variable with the assumption that location on the running variable relative to the treatment threshold perfectly predicts treatment (in this case: being held in EL status). Because districts were allowed to require more than one assessment or allow for more than one pathway to basic skills eligibility, however, we do not anticipate 100% compliance in the probability of being held in EL status below the SBAC threshold, nor 100% compliance in the probability of being reclassified above the SBAC threshold. Consequentially, we use a 2-Stage Least Squares (2SLS) modeling method to evaluate how compliance changes at the threshold of the SBAC cutoff. Our second-stage results are analogous to the treatment-on-treated (TOT) analysis, using only the variation in probability of being held in EL status that is explained by the SBAC cut-off.

Outcomes. Our first set of outcomes relate to standardized test performance on state ELA and math scores as measured by SBAC. We have four academic outcomes: (1) standardized and centered SBAC ELA score by grade and academic year, (2) standardized and centered SBAC math score by grade and academic year; (3) probability of reaching "standard met" (proficiency level 3) or above in ELA; (4) probability of reaching "standard met" (proficiency level 3) or above in math.

We also include two behavioral outcomes. The first is the probability of receiving a suspension over the course of the year and the second is the probability of being chronically absent (defined as missing 10 or more days in the school year). We attempted to include two additional behavioral outcomes related to graduation and expulsion but did not have sufficient sample size or power to run these analyses.

Heterogeneity Analysis Variables. We were also interested in whether effects of reclassification for students who pass their district's ELP threshold but are on the margin of their basic skills threshold varied. Specifically, as outlined in research questions 3 and 4, we are interested in variation by the following factors.

Grade-band. Reclassification may have different effects based on the instructional context that students are in. One foundational predictor of instructional context is schooling level since elementary school students (in grades 5 and below) are generally in a single classroom all day while secondary students (in grades 6 through 12) move between classes. We therefore grouped students by elementary (reclassified at the end of grades 5 or before) and secondary grade band (reclassified at the end of grades 6 or above).

Basic skills criterion difficulty. We categorized districts by the level of difficulty of their basic skills criteria. The easiest or least strenuous category were "SBAC or" districts. These districts provided students with more than one option for meeting the basic skills criterion. For example, a student could meet a set threshold on the SBAC or the i-ready assessment. The middle difficulty level was the "SBAC only" districts. As the name implies, these districts required that students meet the district SBAC basic skills threshold and did not involve any other assessments. The final category was "SBAC and" districts. These districts required that students meet the SBAC basic skills threshold and required them to meet one or more additional assessments in order to be eligible for reclassification.

Basic skills threshold difficulty. We also categorized districts based on the proficiency level that they required on the SBAC for passage of the basic skills criterion. All districts fell into one of three generated categories: (1) proficiency level 3 "standard met", (2) proficiency level 2 "standard nearly met", or (3) somewhere between level 2 and level 3.

Initially, we planned to evaluate whether reclassification effects for students at the margin of the basic skills criterion varied by where districts set their ELP levels, akin to the Robinson-Cimpian and

Thompson (2016) study. However, as will be discussed more below, there was not sufficient variation in ELP thresholds across districts for us to examine this.

Covariates. Our analyses accounted for important differences in student characteristics. These included grade, gender, race/ethnicity (dummy coded for the two most prevalent groups in sample [Latine and Asian] with a third category for all others), a flag for low-socioeconomic status (proxied by FRPL eligibility), home language (dummy coded for the two most prevalent groups in sample [Spanish, Chinese[7]] with a third category for all others), a flag for newcomer status, a flag for special education status, and prior year ELP scale score. Because of a short panel and since our preferred model had district fixed effects, we did not include district-level covariates in any of our models although our descriptive tables include some district level features.

Method

Our first two research questions are descriptive explorations of where our sample of districts set their basic skills criteria and criteria passage patterns among students. To answer these questions, we use frequency counts, percentage calculations, and cross-tabulations. To protect student anonymity and in compliance with our Data Use Agreement, we omit counts and percentages for any group whose cell size was fewer than ten observations.

Our third and fourth research questions employ a fuzzy frontier regression discontinuity (RD) design implemented using the *rdrobust* package in Stata. RD is a quasi-experimental design that, when correct assumptions are met, gives unbiased causal estimates of the impact of a given treatment for a specific population. In our case the treatment is being held in EL status rather than being exited. The population is English proficient students (who meet the state ELP reclassification criterion) and who score close to the basic skills SBAC criterion (N = 14,455 student-year observations; see column 3 of Table 2). The logic behind RD is that English proficient students who score just above or just below the relevant SBAC threshold are, in expectation, identical. One student perhaps guessed correctly on one question and managed to eke past the reclassification criterion, while another guessed wrong and was not eligible to be reclassified. If we compare these two groups of students one year after taking the

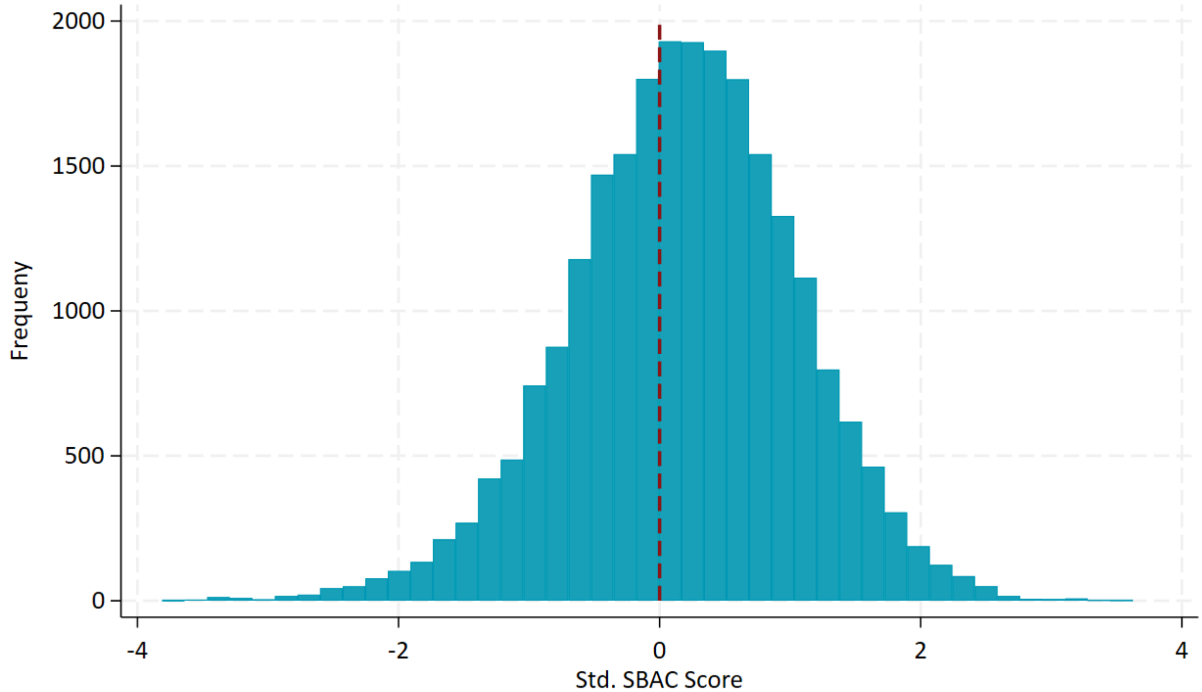
[7] The Chinese language category includes both Mandarin and Cantonese-speaking students.

SBAC test, we can attribute differences in their outcomes to the fact that one group was held in EL status while the other was reclassified (Imbens & Lemieux, 2008; Reardon & Robinson, 2012).

Of note, because we do not have data on students' eligibility based on the teacher evaluation or parental consultation criteria, we cannot include these in our analyses. The omission of these criteria does not invalidate causal inference because variation in eligibility on both criteria should be distributed, like other characteristics, as-good-as random for students scoring just above and just below the academic criterion threshold. But the omission of these two criteria does contribute to our need to conduct a fuzzy rather than a sharp RD. This is because students who just clear the academic threshold (and who we observe above the ELP threshold) may not, for example, have been eligible to be reclassified due to failure to meet teacher evaluation expectations. As such, we do not assume that meeting the academic criterion will necessarily result in reclassification, even among those who meet the English proficiency criterion.

Unbiased causal inference in RD rests on several assumptions. First, there must be no manipulation of the running variable meaning that no one manipulated whether a student scored above or below the basic skills threshold on the SBAC. We tested this by plotting a histogram of the running variable and saw no visible evidence of bunching above or below the centered threshold. We also conducted the McCrary density test and found a non-significant p-value (Figure 2). Note that SBAC scaled scores are discrete rather than continuous; however, following standard guidance, we treat them as continuous because the number of possible values is large (Cattaneo & Titiunik, 2022). In our data set, there are more than 5,000 distinct SBAC scores overall and over 1,000 unique scores within the analytic sample.

Figure 2: Distribution of Prior Year SBAC Scores Among Students who Met District ELP Criteria



Note. SBAC = Smarter Balanced Assessment Consortium. ELP = English language proficiency. Std. = standardized. Visually, we don't see any evidence of manipulation of scores at the test-threshold. We also conducted the formal McCrary manipulation test and found a non-significant estimate ($p = 0.99$).

A second critical assumption is that students who fall just below the reclassification threshold are identical, in expectation, to those who fall just above. The standard way to test this assumption is to check to see if there are measurable discontinuities in pretreatment student covariates at the cut-score. We ran t-tests to assess this assumption and found no statistically significant differences in demographic characteristics, prior English proficiency, or prior math achievement between students scoring below and above the threshold (see appendix Table A5).

In cases where treatment is determined exclusively by a given individual's location on the running variable (i.e. compliance between treatment exposure and treatment eligibility is 100%), researchers can run sharp RDs. In our case we do not expect to see 100% compliance. This is both because prior research suggests that compliance with reclassification policy is low (Johnson, 2020; Mavrogordato & White, 2017), and because we know that districts can include assessments and criteria in their reclassification decisions that we do not observe. Thus, in an "SBAC or" district, for example, we might observe that a student was reclassified that did not meet the SBAC threshold, but this could be because they met an alternative threshold on a non-observed assessment. For this reason, we estimate a 2SLS fuzzy RD. In the first stage, we use an indicator for scoring below the reclassification threshold as an instrument for remaining classified as EL in the following year. In the second stage, we regress our outcomes of interest on the instrumented EL status. This approach identifies the causal effect of being retained in EL status for students whose classification is determined by marginally failing to meet the reclassification criteria—that is, the local average treatment effect induced by the threshold.

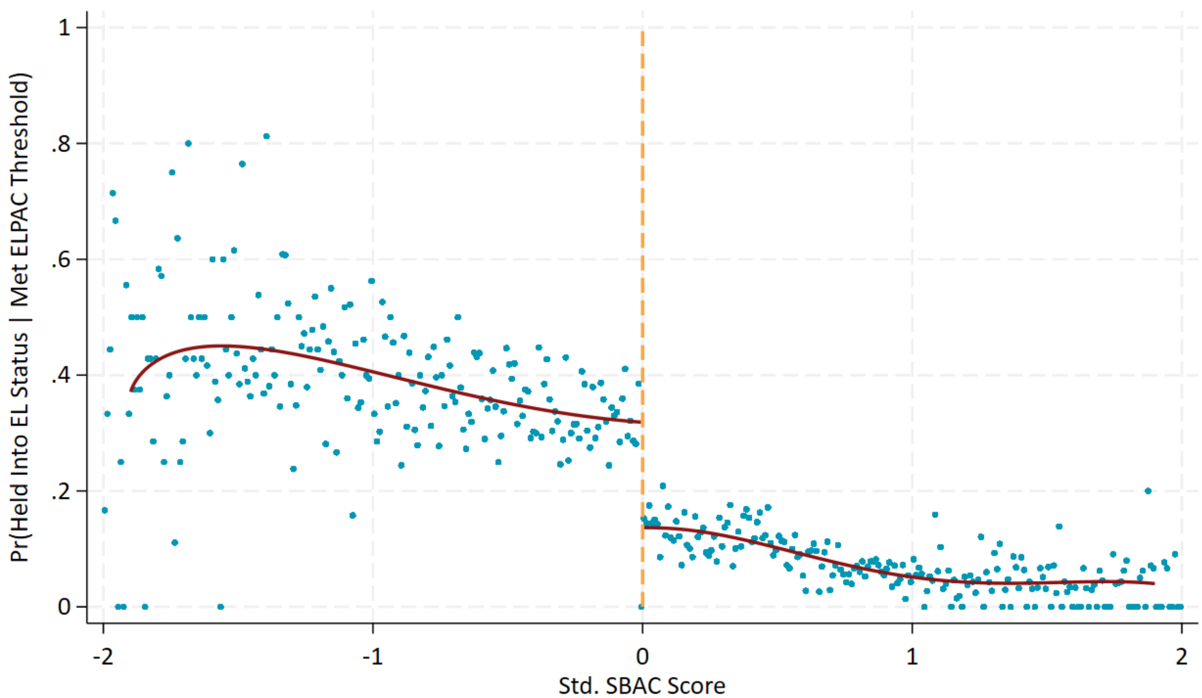
A necessary assumption for the use of a 2SLS RD design is that treatment eligibility meaningfully predicts treatment exposure. A common method of testing this assumption is examining the F-statistics from the first-stage regression models. As such, we tested that SBAC scores predicted retaining EL status using visual inspection and the F-statistics from the first-stage regression models in the mean squared error (MSE) optimal bandwidth. We took an F-statistic >= 10 as evidence for a relevant instrument, as is typically used in education research (Staiger & Stock, 1997). Specifically, for the full sample and each subsample for the heterogeneity analysis, we ran the following regression model (Equation 1), using robust and district-clustered standard errors.

Equation 1

$$EL\_Status_{it} = \pi_0 + \pi_1 * \left(Std.\ SBAC_{it-1}\right) + \pi_2 * I\left(Std.SBAC_{it-1} \leq 0\right) + \pi_3 * I\left(Std.SBAC_{it-1} \leq 0\right) * \left(Std.SBAC_{it-1}\right) + \nu_{it}$$

where π_2 shows the average change in the probability of *EL_Status* for student *i* in year *t* at the threshold of the prior year (*t*-1) SBAC cutoff among students who fall below the threshold on the standardized SBAC score ($Std.SBAC \leq 0$). These results are presented in Table 3 (standard errors are presented in appendix Table A6). Figure 3 displays first stage results for the full sample graphically. As predicted, compliance was quite low in our main sample and across our subsamples of interest. Across bandwidths, eligibility decreased the probability of retention in EL-status in our full sample by between 16 and 17 percentage points. As shown in Figure 3, this low compliance is driven primarily by the fact that many students below the SBAC threshold were reclassified (66%). This is not surprising given that, as will be described soon, most districts allow for non-SBAC assessment scores as alternative routes to passing the basic skills requirement. The change in EL-retention status was far larger, as expected, in the group of "SBAC only" districts (47 to 54 percentage point jump at the threshold, depending on bandwidth). Figure 3 also shows that more than 10% of students in our English-proficient sample who cleared their district's SBAC threshold continued to be observed in EL status the subsequent year. These cases of non-compliance could be (1) due to administrative error (Bartlett et al., 2024; Estrada & Wang, 2015), (2) cases where students did not meet one or both of the parent consultation or teacher evaluation criteria, or (3) where students missed one or more required academic assessments in an "SBAC and" district.

Figure 3: Change in Probability of EL-Retention among Students who Met District ELP Criteria at the Margin of Basic Skills Criterion



Note. EL = English learner. ELP = English language proficiency. ELPAC = English Language Proficiency Assessment for California. SBAC = Smarter Balanced Assessment Consortium.

Table 3: Estimate of the Effect of Academic Criterion Threshold on Probability of EL-status Retention among Students who Met District ELPAC Criteria across Sample Definitions and Bandwidth

Sample Definition	0.25 SD	N	0.5 SD	N	0.75 SD	N	1 SD	N	First-Stage F-Statistics
Full Sample	-0.162**	5337	-0.174***	10223	-0.169***	14455	-0.166**	17617	24.13
SBAC Or	-0.202***	3040	-0.196***	5840	-0.180***	8321	-0.174***	10233	23.32
SBAC Only	-0.470**	393	-0.499***	755	-0.535***	1072	-0.531***	1264	13.44
SBAC And	-0.202	377	-0.212	729	-0.202	1026	-0.197	1277	14.37
SBAC PL2	-0.260***	1830	-0.226***	3559	-0.200**	5129	-0.190**	6408	23.72
SBAC PL3	-0.182~	1980	-0.216*	3765	-0.223*	5290	-0.222*	6366	11.64
SBAC PL Other	-0.029	6193	-0.064	6193	-0.066	6193	-0.068	6193	8.46
ELPAC Or	-0.159**	5016	-0.175***	9578	-0.174**	13542	-0.172**	16492	24.38
ELPAC And	-0.188	307	-0.153	620	-0.083	880	-0.061	1078	2.28
Elementary grades	-0.148**	2578	-0.153**	4951	-0.153**	7039	-0.147**	8670	15.97
Secondary grades	-0.172**	2759	-0.193**	5272	-0.186**	7416	-0.183**	8947	26.59

Note. SD = standard deviation. SBAC = Smarter balanced assessment consortium. ELPAC = English language proficiency assessment for California. Models estimated by regressing the flag of entering current year as EL on standardized and centered version of past year SBAC score, flag for scoring below the cutoff, and an interaction between the two terms. Point estimates show the change in average EL-retention probability at the margin of the SBAC threshold. Inference based on heteroskedastic-robust and clustered standard errors. Std. errors for the estimates are presented in appendix table A6.
p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

As a result of this analysis, we removed two groups from our subgroup analyses (research questions 3b and 4). The first was the SBAC proficiency threshold "something else" (i.e., between 2 and 3; F-statistic = 8.46).[8] Second, we had initially included an additional research question related to whether reclassification effects varied at the margin of the basic skills thresholds for students in districts with high versus low ELP reclassification criteria. We had to drop this question, however, because the F-statistic for non-typical ELP districts (those whose ELP criteria was not reaching a 4 ["proficient"] on the overall ELPAC) was 2.28, far below the cutoff of 10. Table A1 in the appendix provides descriptive statistics of the sample, divided by ELP threshold.

[8] We kept these districts in our full sample analyses but removed them from our subgroup analyses. So, for example, there are no districts with SBAC proficiency thresholds between 2 and 3 even in our subgroups related to SBAC 'and', 'or', and 'only.'

Because we are interested in how the academic criterion for reclassification impacts students in California, we employ a frontier RD design. This means that rather than including students close to any of the reclassification criteria thresholds, we limit our sample to students who clear the ELP criterion and are therefore considered, by their district and the state, to be English proficient. Frontier RDs allow for a more targeted understanding of the implications of a specific criterion (in our case the academic criterion) and have the added benefit of reducing a multi-factored eligibility context into a single factor (Reardon & Robinson, 2012).

Before estimating our 2SLS model we first ran Intent-to-treat (ITT) models (i.e. sharp RDs). Estimates from these models should be interpreted as the effect of not meeting the reclassification criteria on student outcomes. These ITT models use the following general equation:

Equation 2

$$Y_{it} = \beta_1 * I\left(Std\,SBAC_{it-1} \leq 0\right) + \beta_2 * f\left(Std\,Score_{it-1}\right) + \beta_3 * I\left(Std\,SBAC_{it-1} \leq 0\right) * f\left(Std\,Score_{it-1}\right) + X_{it} + T_t + D_i + \epsilon_{it}$$

where, Y_{it} is a generic outcome of interest for student *i* in year t. $I(Std.\ SBAC \geq 0)_{it-1}$ is a flag for if student i met the district-set SBAC threshold for reclassification in the prior Spring administration of SBAC. The running variable, $Std\,Score_{it-1}$, is a student's score and $f(.)$ is a flexible function. X_{it} is a vector of student-level covariates including demographic characteristics, T_t is a set of year-fixed effects, and D_i is a set of district-fixed effects. The parameter of interest, β_1, represents the effect of not meeting the SBAC threshold on students' outcomes for English proficient students near the threshold of SBAC eligibility. Our models employed triangular kernels, local linear fits, and robust and clustered standard errors for all inferences.

Across our regression discontinuity models we evaluate the statistical significance, direction, and magnitude of these effect estimates, using Kraft's (2020) benchmarks of effect size in educational interventions. We consider it useful to interpret direction and magnitude in addition to statistical significance because our ability to detect statistically significant effects may be limited. Non-significant estimates may reflect true null effects of reclassification, but they may also arise from other factors such as examining outcomes only one year after reclassification (if effects accrue over time) or large standard errors due to SBAC scores serving as a relatively weak instrument for reclassification. Consistent with guidance that scientific conclusions should not rely solely on statistical significance (Gelman & Stern, 2006; Wasserstein & Lazar, 2016), we interpret estimates of meaningful magnitude or policy relevance as potentially substantively important while emphasizing the considerable uncertainty surrounding them and the need for further research.

After running our ITT models, we moved on to our fuzzy RD models. These are treatment-on-treated (TOT) models which estimate the effect of being held in EL status on student outcomes. This TOT model uses the following general equation:

Equation 3

Stage 1:

$$EL\_Status_{it} = \pi_1 * g(Std.\ SBAC_{it-1}) + \pi_2 * I(Std.\ SBAC_{it-1} \leq 0) + \pi_3 * I(Std.\ SBAC \leq 0)_{it-1} * g(Std.\ SBAC_{it-1}) + X_{it} + T_t + D_i + \nu_{it}$$

Stage 2:

$$Y_{it} = \beta_1 * EL\_\hat{Status}_{it} + \beta_2 * h(Std.\ Score_{it-1}) + \beta_3 * h(Std.\ Score_{it-1}) * EL\_\hat{Status}_{it} + X_{it} + T_t + D_i + \epsilon_{it}$$

The parameter of interest, β_1, represents the effect of being held in EL status because of missing the SBAC cutoff (among students who met the ELP criterion) on next-year student outcomes. All other variables have the same meaning as above with the addition of EL_Status_{it} which represents a dummy variable for students retained in EL-status. We again used triangular kernel, linear fits, and robust and clustered standard errors for all inferences.

Importantly, as in all RD designs, the population of causal inference is limited to students near the treatment threshold. In this analysis, this relates to students who meet district ELP thresholds (proficiency level 4 for nearly all students) and are at the margin of district-specific basic skills criteria. We cannot assume that the impact of retaining EL-status would be the same for students who did not meet the English proficiency criterion nor for students who score far above or far below their district's basic skills SBAC threshold. Causal inference is further limited to the districts and grades that we examined. We ran all ITT and TOT models for multiple bandwidths, including mean-square-error optimal bandwidth for each outcome and specification.

The above models answered research question 3 which asks about the main effect of being held in EL status for English proficient students at the margin of basic skills eligibility. We also conducted several robustness checks for the main effects including adjusting for mass points, using uniform kernels, fitting local quadratic polynomials, using only robust standard errors, and testing placebo cutoffs (results, shared in the appendix, are described in the results section). The second part of research question 3 and research question 4 ask about heterogeneity in the effects of reclassification at the academic margin. Specifically, in research question 3 we look to see if reclassification effects differ by grade band, separating our sample of students into elementary (those who were reclassified based on 3rd, 4th, or 5th grade SBAC eligibility) and secondary (those who were reclassified based on , 6th, 7th, 8th, or 11th grade SBAC eligibility).

To answer this question, we ran separate models by subgroup and then computed difference-in-regression discontinuity (DiRD) estimates, following the approach used by Bartlett et al. (2026). Separate models followed the same specifications shown in Equations 1 and 2 but the sample of each was limited to a specific subgroup (e.g. elementary students). The DiRD approach estimates the magnitude and statistical significance of the difference between effects across two subgroups to examine the extent to which effects of being held in EL status vary across subgroups. To find the DiRD estimate, we subtracted the subgroup-specific estimates:

Equation 4a

$$DiRD = \beta_{Subgroup\ i} - \beta_{Subgroup\ j}$$

After obtaining a DiRD estimate of the difference between groups, we computed standard errors for each DiRD estimate using the equation:

Equation 4b

$$SE(DiRD) = \sqrt{var(\beta_{Subgroup\ i}) + var(\beta_{Subgroup\ j})}$$

Here, *i* and *j* represents two groups, for example students in elementary grades and students in secondary grades.

Research question 4 asks about how, if at all, the effect of retaining English-proficient students in EL status at the academic margin differs based on the difficulty level of the basic skills criterion. Similar to our analysis by grade band, we ran both separate models by difficulty subgroup, and calculated a DiRD estimate. As a reminder, we ran two sets of analyses in research question 4, first exploring variation by criterion difficulty ("SBAC or", "SBAC only", and "SBAC and") and then by SBAC threshold difficulty (proficiency level 2, proficiency level 3). In analyses for RQ3 and RQ4, we retained only covariates whose categories contained more than ten observations, eliminating the need for additional adjustments.

Finally, research question 5 asks about how, if at all, the effect of retaining English-proficient students in EL-status at the academic margin differed across districts with similar difficulty levels of the basic skills criterion. Following the same ITT approach shown in Equation 2, we estimated separate unadjusted models for each district that had at least five students on either side of the threshold within the optimal mean square error bandwidth.

Equation 5

$$Y_{itd} = \beta_1 * I\left(Std\, SBAC_{idt-1} \leq 0\right) + \beta_2 * f\left(Std\, Score_{idt-1}\right) + \beta_3 * I\left(Std\, SBAC_{idt-1} \leq 0\right) * f\left(Std\, Score_{idt-1}\right) + \epsilon_{itd}$$

β_1 in these models represents the district (*d*) specific estimates of being held in EL status. As above, we used local linear fits, triangular kernels, and robust standard errors. Then, we plotted these district-specific estimates in forest plots to visualize between-district variation. Because many of our within-district sample sizes are small and compliance with reclassification criteria low, we expect many of these estimates to be imprecise. However, to guard against very low compliance, we report district-specific estimates for districts meeting one of the two following criteria: a) first-stage F-statistics is greater than or equal to 10, or, b) a statistically significant change in the probability of retention into EL status at the district-specific basic skills threshold ($\alpha = 0.05$). We consider this analysis exploratory.

Results

Research Question 1: What Are the Basic Skills and ELP Reclassification Criteria in A Set of Randomly Sampled California Districts?

Basic Skills Criterion

Our first question explores how school districts in California set their basic skills and ELP reclassification criteria. California law stipulates that districts set their own basic skills reclassification criteria in such a way that it provides a point of comparison of the individual student's English language arts "basic skills" against the performance level of same-age English proficient peers (California Education Code § 313(f), 2025). Statute does not provide much additional information, however, leaving it up to individual districts to select what assessment or assessments to use, who the comparison group of same-age peers should be (within district, within state, etc.), or how to set the reclassification cut score on that assessment or assessments, for example. As a result, it is important to look at how districts have set this criterion since where and how a district sets their basic skills criterion influences who is reclassified and when. It is also worth noting that districts' basic skills criteria likely reflect both conscious decisions on the part of the district (i.e. what level of difficulty they want) as well as less-purposeful decisions that may reflect confusion or lack of capacity at the district level to make this high-stakes determination (Mercado-Garcia et al., 2026).

As described in the method section above, we limited our sample to districts that used the SBAC in their basic skills criterion. This was a necessary sample restriction in order to run our regression discontinuity models. Nearly one in five sampled districts (amongst those for which we could find reclassification criteria) did not use the SBAC. While these non-SBAC districts were dropped from our sample, it is noteworthy that a considerable proportion of California districts do not use the state standardized ELA assessment as part of their basic skills criterion.

Our final sample of 76 districts that used SBAC in their basic skills reclassification criterion varied widely both in terms of whether they paired SBAC with any other ELA assessments and where on the SBAC they set their thresholds (see Table 4).

Table 4: Characteristics of Reclassification Criteria among Sampled Districts

	# Students-Year Observations	# District-Grade Combinations
Total	112,077	430 (76 districts)
ELPAC proficiency level threshold		
3.5	123 (0.1%)	5 (1.1%)
4	111,954 (99.9%)	425(98.9%)
ELP criterion difficulty level		
Overall ELPAC or	687 (0.6%)	7 (1.6%)
Overall ELPAC only	104,528 (93.3%)	381 (88.6%)
Overall ELPAC and	6,862 (6.1%)	42 (9.8%)
Basic skills criterion difficulty level		
SBAC or	95,433 (85.1%)	344 (80.0%)
SBAC only	10,845 (9.7%)	61 (14.2%)
SBAC and	5,799 (5.2%)	25 (5.8%)
District SBAC proficiency level threshold		
2	45,154 (40.3%)	138 (32.1%)
Between 2-3	31,293 (27.9%)	98 (22.8%)
3	35,630 (31.8%)	194 (45.1%)

Note. ELP = English language proficiency. ELPAC = English language proficiency assessment for California. SBAC = Smarter balanced assessment consortium. “Or” districts are districts that require that students meet a given threshold on the named assessment (ELPAC for English language proficiency [ELP] criterion, SBAC for basic skills criterion) or a threshold on one or more alternative assessments. “Only” districts require only the named assessment. “And” districts require that students meet the threshold on the named assessment and meet one or more additional thresholds on other assessments or assessment components.

Pairing SBAC with other assessments. While not explicitly mentioned in state education code, we found that districts varied greatly in terms of whether they allowed one or more non-SBAC assessments to replace the SBAC ELA assessment as the basic skills criterion (“SBAC or” districts), whether they required any assessments in addition to SBAC (“SBAC and” districts), or whether they required only the SBAC (“SBAC only” districts).

By far, the largest group of districts (80.2%) were “SBAC or” districts, providing students with multiple options for clearing the basic skills criterion. 85.3% of sampled students fell into this category. 14.2% of districts (serving 9.7% of students) were “SBAC only” while 5.8% were “SBAC and” districts (serving 5.2% of the student sample), requiring multiple ELA thresholds to meet the basic skills criterion. Among the "SBAC or" and "SBAC and" districts, there was an array of alternative assessments. Common assessments included the i-Ready, Star Reading, and MAP Growth assessments. We consider "SBAC or" districts to have more lenient basic skills criteria since students in these districts had multiple chances to meet the basic skills criterion, and multiple opportunities during the school year to clear the threshold. "SBAC only" and "SBAC and" districts are both more strenuous. "SBAC only" districts, for example, put the weight of the basic skills criterion on a single test score taken once per year. "SBAC and" are even more strenuous because they require that single high-stakes assessment score and additional assessment-based thresholds.

SBAC proficiency level. Districts also varied with regard to the threshold they required on the SBAC ELA assessment. Districts fell into one of three, somewhat equivalently-sized categories. The largest group of districts was those that set their proficiency level at "standard met" (level 3) comprising 45.1% of our sample of districts. These were, on average, smaller districts, since they served only 31.8% of the student sample. The second largest set of districts was those that set their threshold at 2 or "standard nearly met" comprising 32.1% of districts serving 40.3% of students. Finally, a somewhat smaller group of districts (22.8%; serving 27.9% of students) set their threshold somewhere between 2 and 3. No districts set their threshold above 3 or below 2.

State statute specifies that the basic skills criterion should be set relative to the performance of same-age English proficient peers. At the state level, the average SBAC ELA proficiency level of non-ELs is 2.53, indicating that the 32.1% of districts that have set level 2 may be considerably under that average, while the 45.1% that set a level 3 are considerably above the state mean. An alternate comparison group, however, may not be the full state, but instead be English proficient students who are similar to EL-classified students (Faulkner-Bond, 2026). For example, districts might choose to set the comparison group to English proficient peers within the district. Alternatively, 84.4% of ELs in California are Latine, so an appropriate comparison group (and therefore threshold) might the be average score of non-EL Latine students in the state, which is 2.27. Similarly, more than three-quarters (78.5%) of ELs live in low-income households that are eligible for free or reduced-price lunch (FRPL). Thus, an appropriate comparison group might, instead, be non-EL FRPL-eligible students. Statewide, their average SBAC score is 2.17. Taken together, we conclude that an appropriate threshold, given the terms of state education code, is likely between SBAC proficiency level 2.2 and 2.5. We find, then that about a third of districts are set below that threshold (2), a third above (3), and a third are relatively close to that threshold range.

English Language Proficiency Criterion

Unlike California's basic skills criterion, the ELP criterion is clearly set at the state level. State statute specifies ELPAC level 4 on the summative assessment as the "statewide standardized ELP criterion" (California Department of Education, 2025) with flexibility only allowable for students with disabilities and these departures being clearly spelled out. As such, we expected to see either no variation, or far less variation, in the ELP criteria set forth in district reclassification guidelines.

As hypothesized, we found that 98.9% of our sampled districts set their district ELP threshold at ELPAC level 4 (see Table 4). These districts accounted for 99.9% of sampled students. Likewise, over 90% of districts had an "overall ELPAC only" policy meaning that students cleared the ELP reclassification threshold when they hit the ELPAC overall proficiency level specified. About 10% of districts had an "overall ELPAC and" rule, usually indicating that they required an overall ELPAC score of 4 but also one or more of the ELPAC's sub-scores for reading, writing, speaking, or listening. A very small number of district observations had an "overall ELPAC or" policy (easy; 1.6%). Table A2 in the appendix shows descriptive statistics by ELPAC criterion type.

Research Question 2: Among Our Sample of Districts, How Many Students Met Reclassification Criteria for Basic Skills Alone, ELP Alone, Both Basic Skills and ELP, and How Many Were Subsequently Reclassified? How Does this Vary by Subgroup?

Our second research question asked about how students were faring on the basic skills criterion in their district, how that compared to student outcomes on the ELP criterion, and the extent to which reclassification outcomes appeared to align with eligibility on these two assessments. As a reminder, because over 85% of districts allowed for alternative assessments to replace the SBAC, we expected to find that more students were reclassified than passed the SBAC threshold, so long as they met the ELP criterion.

Table 5 shows the proportion of students, by characteristic and context, who met various iterations of the basic skills and ELP reclassification criteria and Table 6 shows the proportion of students, by criteria met, student characteristic, context, who were reclassified. A primary finding was that even looking solely at SBAC scores (i.e. not accounting for allowable alternative basic skills assessments) more students met the basic skills criterion than the ELP criterion. Across our sample, 26.6% of students met the basic skills criterion while 21.1% met ELP. (In the appendix, we include similar tables for the overall state; See Tables A3 and A4).

Table 5: Proportion of Students by Subgroup Meeting District Reclassification Criteria and Reclassifying

	District Sample	Met District ELP Criteria*	Met District SBAC Criteria*	Score > District EO Mean on SBAC*	Score > State EO Mean on SBAC*	Met District ELP and SBAC Criteria	Entered Next Grade Reclassified
		Row percent	Row percent	Row percent	Row percent	Row percent	Row percent
N	112,077	21.1	26.6	17.6	14.9	12.7	21.8
Reclass							
No	87,656	5.3	17.3	10.5	7.5	1.4	
Yes	24,421	77.8	59.9	43.3	41.2	53.2	
Current Gr*							
4	21,330	13.4	33.3	26.0	23.2	11.2	15.8
5	21,379	20.2	30.0	23.0	20.2	14.6	22.8
6	19,236	24.8	27.8	18.5	14.8	15.3	24.6
7	16,523	21.7	23.6	14.2	11.4	11.7	22.1
8	14,740	28.5	23.4	13.9	10.3	14.3	28.1
9	9,252	27.1	18.7	9.1	7.3	10.9	21.1
12	9,503	15.3	19.0	5.5	4.9	7.4	17.7
Past-Yr Tested Gr*							
3	21,337	13.4	33.3	26.0	23.2	11.2	15.8
4	21,373	20.2	30.0	23.0	20.2	14.6	22.8
5	19,240	24.8	27.8	18.5	14.8	15.3	24.6
6	16,525	21.7	23.6	14.2	11.4	11.7	22.1
7	14,745	28.5	23.4	13.9	10.3	14.3	28.1
8	9,253	27.1	18.7	9.1	7.3	10.9	21.1
11	9,604	15.2	18.9	5.5	4.9	7.3	17.6
Aca Yr							
2022-2023	57,120	20.6	27.7	18.3	15.6	12.9	20.9
2023-2024	54,957	21.7	25.4	17.0	14.1	12.5	22.8
Gender							
Female	51,690	20.9	29.0	19.6	16.3	13.2	22.1
Male	60,377	21.3	24.5	16.0	13.7	12.2	21.6
Low Socio-Economic Status Flag							
No	15,615	30.4	38.0	21.1	26.5	21.2	27.9
Yes	96,462	19.6	24.7	17.1	13.0	11.3	20.8
Race/Eth							
Black	513	25.1	32.2	19.5	18.1	16.8	22.0
Asian	15,44	31.6	42.1	24.0	30.6	23.1	28.4
Filipino	1,180	25.8	44.7	23.1	25.6	19.3	25.4
Hispanic	88,99	18.9	23.2	16.5	11.6	10.4	20.4
Other	1,908	24.1	30.3	16.5	17.6	15.6	22.6
White	4,032	28.0	33.7	17.7	22.9	18.4	25.4
Language							
Arabic	1,764	23.8	26.4	15.0	18.1	14.1	22.8
Chinese	3,010	38.6	52.5	30.0	39.9	30.6	35.9
Other	14,376	27.3	37.5	19.8	24.0	19.1	24.1
Russian	592	37.7	40.2	21.6	30.9	25.8	32.9

	District Sample	Met District ELP Criteria*	Met District SBAC Criteria*	Score > District EO Mean on SBAC*	Score > State EO Mean on SBAC*	Met District ELP and SBAC Criteria	Entered Next Grade Reclassified
Spanish	88,833	18.9	23.2	16.5	11.5	10.4	20.4
Vietnamese	3,502	34.2	43.6	26.9	36.0	25.2	31.8
Special Ed**							
No	90,620	24.5	30.4	20.7	17.3	14.9	23.7
Yes	21,456	7.1	10.2	4.7	4.5	3.2	13.7
Years in EL							
<=2	2,372	10.8	17.8	9.4	12.0	8.0	10.8
3--5	31,71	18.2	30.4	22.5	20.5	13.4	20.0
5+	77,991	22.7	25.3	15.9	12.7	12.5	22.8
Newcomer							
No	104,236	22.1	27.4	18.4	15.3	13.1	22.7
Yes	7,841	8.8	14.9	7.8	9.7	6.5	9.4
Past Gr Math							
1	75,35	12.6	14.4	8.2	4.5	5.2	13.7
2	23,43	31.7	42.7	29.4	24.4	20.3	32.1
3	9,183	46.2	64.2	48.2	51.7	36.5	45.9
4	3,618	67.1	81.2	63.0	77.1	60.0	63.3
Missing	265	11.7	10.9	6.0	4.5	5.7	12.1
District Characteristics							
District Locale							
City	53,736	22.0	25.0	16.4	15.8	13.0	21.1
Suburb	42,210	20.7	29.0	18.3	14.8	12.9	22.2
Town	13,899	19.8	27.0	19.9	11.8	11.7	23.8
Rural	2,232	18.1	13.4	20.7	13.1	6.6	18.8
Basic skills criterion difficulty level							
SBAC or	95,433	20.4	26.4	17.6	14.1	12.3	21.7
SBAC only	10,845	23.3	26.8	19.1	17.9	13.5	20.7
SBAC and	5,799	29.1	28.7	15.3	21.6	17.2	25.8
District SBAC proficiency level threshold							
2	45,154	20.4	36.7	18.1	13.5	15.3	21.7
Between 2 and 3	31,293	19.8	22.4	20.5	12.8	10.7	20.7
3	35,630	23.3	17.4	14.5	18.4	11.0	22.9

Note. Gr = grade. Aca = academic. Yr = year. Eth = ethnicity. Ed = education. EL = English learner. ELP = English language proficiency. SBAC = Smarter balanced assessment consortium. *We restricted to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past-year. **Among those who take the unmodified SBAC.

Table 6: Proportion of Students Reclassified in Sampled Districts, by Eligibility Status

	Reclassified with			
	Met Only ELPAC	Met Only SBAC	Met Both	Met Neither
Prop.	63.4	10.5	91.5	5.2
Current Grade level*				
3	-	-	-	-
4	75.7	7.0	98.8	3.3
5	71.5	9.7	99.2	5.7
6	65.2	11.1	98.9	4.8
7	61.4	11.4	98.6	6.0
8	60.8	19.2	98.0	7.7
9	57.0	8.0	96.7	3.2
11	-	-	-	-
12	62.7	15.3	93.6	6.0
Past-Year SBAC Grade Tested*				
3	75.7	7.0	98.7	3.3
4	71.4	9.7	99.3	5.7
5	65.3	11.1	98.9	4.8
6	61.4	11.4	98.6	6.0
7	60.8	19.2	98.0	7.7
8	57.0	8.0	97.0	3.2
11	62.4	15.2	93.4	5.9
Academic Year				
2022-2023	60.0	9.8	98.5	4.9
2023-2024	66.3	11.2	98.1	5.5
Gender				
Female	66.0	10.7	92.0	4.9
Male	61.5	10.2	90.9	5.4
Low SES Flag				
No	51.2	9.4	89.7	4.8
Yes	65.6	10.7	92.0	5.2
Race/Ethnicity				
African American	58.1	-	96.6	-
Asian	51.8	7.0	98.4	3.4
Filipino	61.0	5.7	98.7	2.8
Hispanic	65.8	11.7	98.6	5.5
Other	61.1	5.4	97.1	4.1
White	58.1	9.9	95.1	3.7
Language				
Arabic	52.9	15.7	96.1	4.0
Chinese	58.5	9.1	98.2	4.0
Other	51.4	6.1	97.3	3.2
Russian	68.6	-	95.0	3.9
Spanish	65.9	11.7	98.6	5.6
Vietnamese	56.3	6.7	99.7	3.4
Special Education (among those who take unmodified SBAC)				
No	63.4	10.2	98.5	4.0
Yes	62.8	12.9	95.9	8.6

	Reclassified with			
	Met Only ELPAC	Met Only SBAC	Met Both	Met Neither
Years in EL				
<=2	45.5	9.4	99.5	2.1
3-5	68.9	7.5	99.8	4.5
5+	62.5	12.1	97.7	5.6
Newcomer Status				
No	63.7	10.4	98.4	5.5
Yes	47.5	11.2	96.8	2.1
Past Grade Math Performance Level				
1	62.3	8.6	99.1	4.7
2	66.0	11.7	99.7	7.0
3	65.2	11.9	99.2	8.4
4	55.4	13.8	98.8	7.8
Missing	-	-	55.6	5.0
District Characteristics				
District Locale				
City	56.7	10.8	98.3	4.8
Suburb	73.5	7.0	98.0	5.0
Town	60.4	20.7	100.0	7.3
Rural	70.7	8.6	96.7	5.8
Basic skills criterion difficulty level				
SBAC or	67.7	9.79	93.5	5.0
SBAC only	40.5	14.57	87.3	4.7
SBAC and	50.0	15.12	73.6	9.2
District SBAC proficiency level threshold				
2	61.9	7.07	93.6	4.7
Between 2 and 3	65.4	16.24	90.6	4.6
3	62.9	15.61	88.5	6.2

Note. EL = English learner. ELA = English language arts. SES = socioeconomic status. Prop. = proportion. Low-SES refers to students eligible for free- or reduced-price lunch. ELPAC = English language proficiency assessment for California. SBAC = Smarter balanced assessment consortium *We restricted to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past-year. - Indicates that the N size was below 10 and we therefore do not report percentages.

While this pattern held for most subgroups, an important exception was student grade level. Students in elementary were more likely to meet the basic skills criterion than the ELP criterion, at times by a factor of more than two to one. But by 7th grade, students were more likely to meet the ELP than the basic skills criterion (although the pattern reversed again in 11th grade). This is likely because many EL students are in U.S. schools from kindergarten onward and their English improves year by year. At the same time, the SBAC academic ELA assessment becomes increasingly challenging as it aligns with advancing grade level academic standards. As a result, of the two criteria, the basic skills criterion poses the larger hurdle for students in secondary school.

There were two other notable contexts in which fewer students met district SBAC thresholds than ELPAC thresholds. The first was districts with more strenuous basic skills criteria. Fewer students met basic skills criteria in "SBAC and" districts than met the ELPAC threshold. This relationship was even more pronounced in districts that set their SBAC threshold to level 3 (above the state mean). In these districts students were six percentage points more likely to pass ELPAC than SBAC. The final context in which the basic skills criterion was more challenging than the ELP criterion was rural districts. SBAC and ELPAC pass rates were lower overall in rural districts than in cities, towns, and suburban districts, and SBAC pass rates were considerably lower (13.4%) than ELPAC pass rates (18.1%). As we will show later, this is likely partially because rural districts were more likely to have strenuous basic skills criteria.

As a result, there are large numbers of students who demonstrate proficiency on the ELPAC but were not reclassified the subsequent year. We calculate that of the 23,693 students in our sample that scored ELPAC level 4, 4,684 were not reclassified the subsequent year (19.7%). A similar proportion, 18.0%, of students at the state level were in the same position (36,895 cases).[9]

The second to final column in Table 5 illustrates another surprising finding: There was limited alignment between the students who passed the ELPAC and the SBAC criteria. Across the sample, over 20% of students met each of the two main reclassification criteria, but only 12.7% of students met both in the same year. This means that about half of students who met SBAC did not meet ELPAC and vice versa. In other words, a significant proportion of students who demonstrated full English proficiency did not demonstrate academic performance on par with their English proficient peers. Meanwhile a significant proportion of students who demonstrated parity or above-parity academic performance were not considered English proficient.

Table 6 confirms our hypotheses about the relationship of criterion pass rates to reclassification outcomes. Across the full sample 91.5% of students who met district basic skills and ELP thresholds were reclassified the subsequent year while only 5.2% of students who met neither criterion were

[9] These figures can be derived (approximately, due to rounding) from percentages shown in Table 3 (district sample) and Appendix Table A1 (state sample).

reclassified. Adherence to the ELPAC threshold was high, only 10.5% of students who met the SBAC but not the ELPAC threshold were reclassified. Because of the large proportion of districts allowing alternative basic skills assessment to the SBAC, it is not surprising that a full 63.4% of students who met the ELPAC threshold but not the SBAC threshold were reclassified. We expect that many of these students met alternative basic skills assessment thresholds that we are unable to observe in our dataset. This pattern was quite stable across subgroups and contexts.

Research Question 3: Accounting for District Specific Criteria, What is the Impact Being Held in EL Status on Academic and Behavioral Outcomes among Otherwise Eligible Students Who Pass the ELP Criteria and are on the Margin of the Basic Skills Criterion? How Does this Impact Vary by Grade Band?

Across the full sample there were no statistically significant subsequent-year impacts of being held in EL status on academic or behavioral outcomes among students who met the ELP criterion but were on the margin of the basic skills criterion (see Table 7 for results of our preferred models, and Table A7 in the appendix for alternative models and bandwidths). The first set of four columns show results from our preferred ITT model while the second set of four columns show our preferred TOT model results. Our preferred models, for both ITT and TOT, includes district fixed effects (FE) and covariates using an optimal bandwidth, calculated separately for each model. Figures 4 through 9 graphically display the ITT results for each outcome (these are based on our baseline model results which are shown as Model 1 in Appendix Table A7).

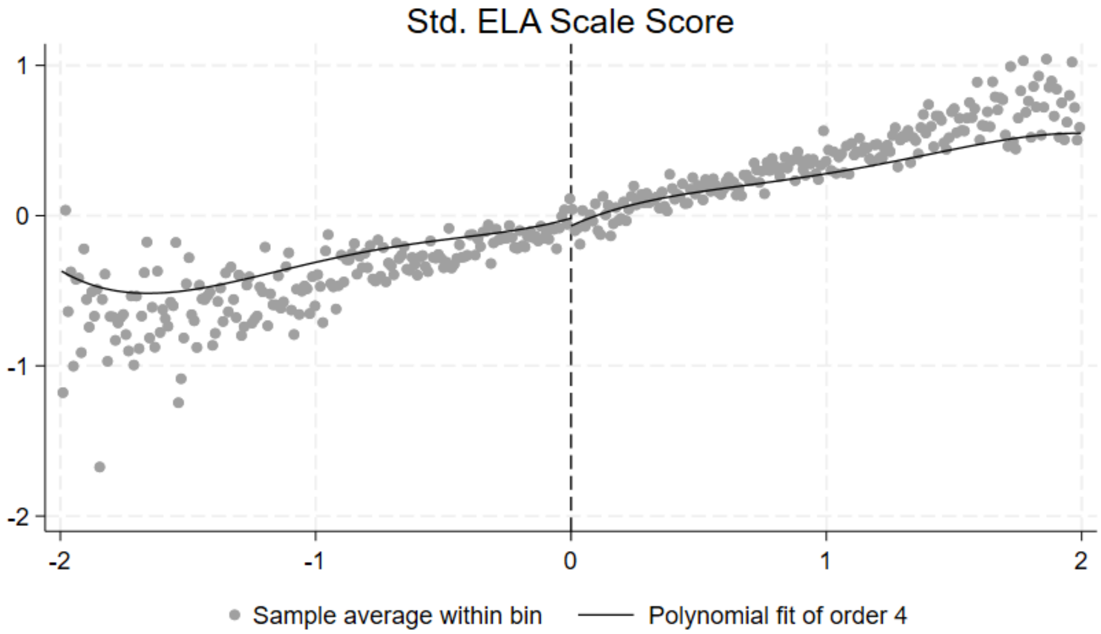
Table 7: Estimates of the Effect of EL-status Retention among Students who Met District ELPAC criteria for Students on the Margin of Basic Skills Criteria across Outcomes, Optimal Bandwidth, and Preferred Model Specification

	ITT				TOT			
Outcome	Optimal BW	N in Optimal BW	Estimate	Std. Error	Optimal BW	N in Optimal BW	Estimate	Std. Error
Std. ELA Score	0.476	7970	0.024	0.043	0.646	10463	0.107	0.244
Std. Math Score	0.806	12493	-0.010	0.048	0.724	11497	-0.041	0.297
Pr(ELA >= 3)	0.648	10495	0.007	0.029	0.975	14219	-0.038	0.167
Pr(Math >= 3)	0.84	12883	-0.010	0.024	0.738	11678	-0.053	0.151
Pr(Suspension)	0.677	13283	0.005	0.007	0.676	13275	0.028	0.046
Pr(Chronic Absenteeism)	0.647	12776	0.032	0.021	0.729	14064	0.186	0.129

Note. ITT = Intent to treat. TOT = Treatment on the treated. ELA = English language arts. PL = Proficiency level. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model). Models estimated using *rdrobust* package in Stata with triangular kernels and linear fits. Estimates show the change in average outcome at the margin of the SBAC threshold. Inference based on heterosckedastic-robust and clustered standard errors. Covariates include flags for race/ethnicity, language code, indicator for low socio-economic status, indicator for special education placement, grade, newcomer status, and past year ELP scaled score. Estimates across other bandwidths and model specification are presented in table A7.

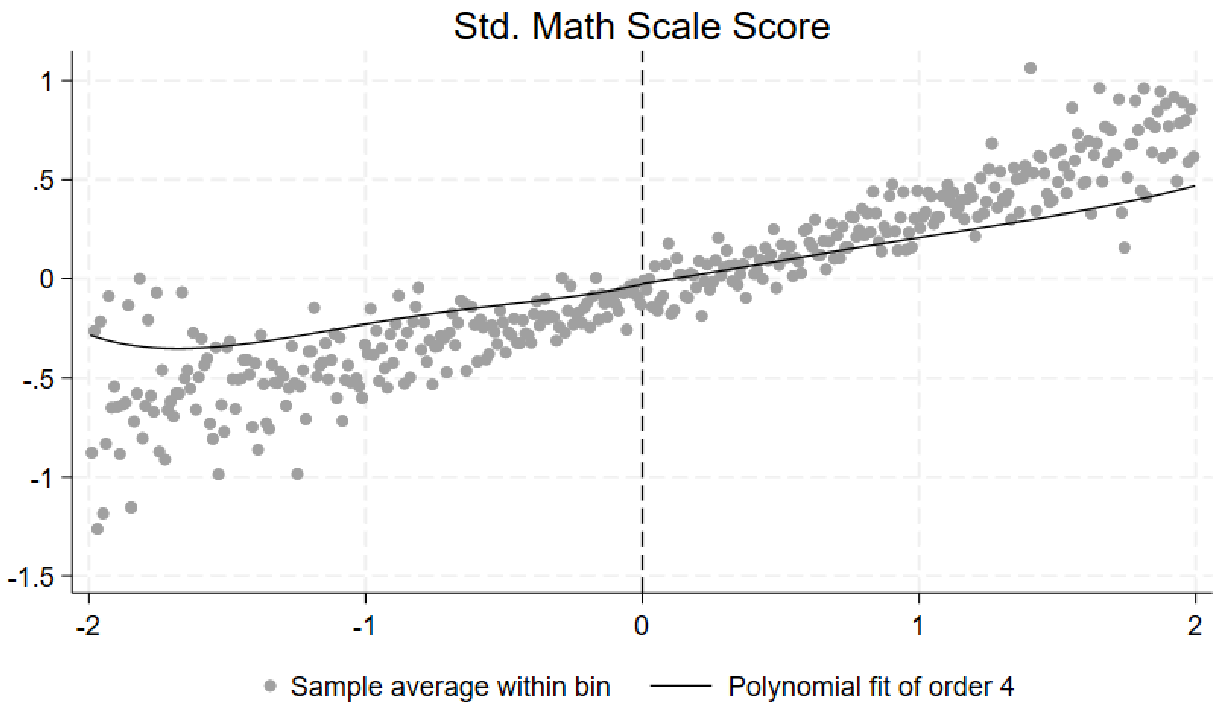
p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

Figure 4: Change in Standardized ELA Score at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



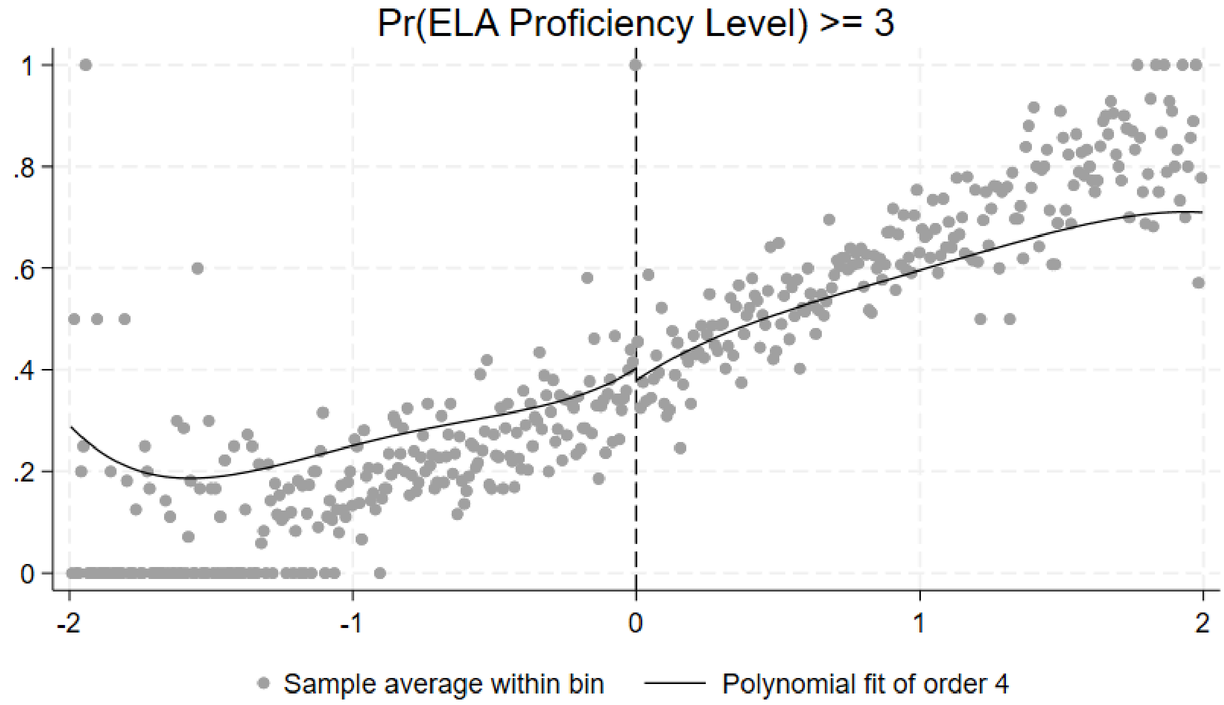
Note. ELA = English language arts. Std. = standardized. The figure is adjusted for student characteristics and district fixed effect using `rdrobust` Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 5: Change in Standardized Math Score at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



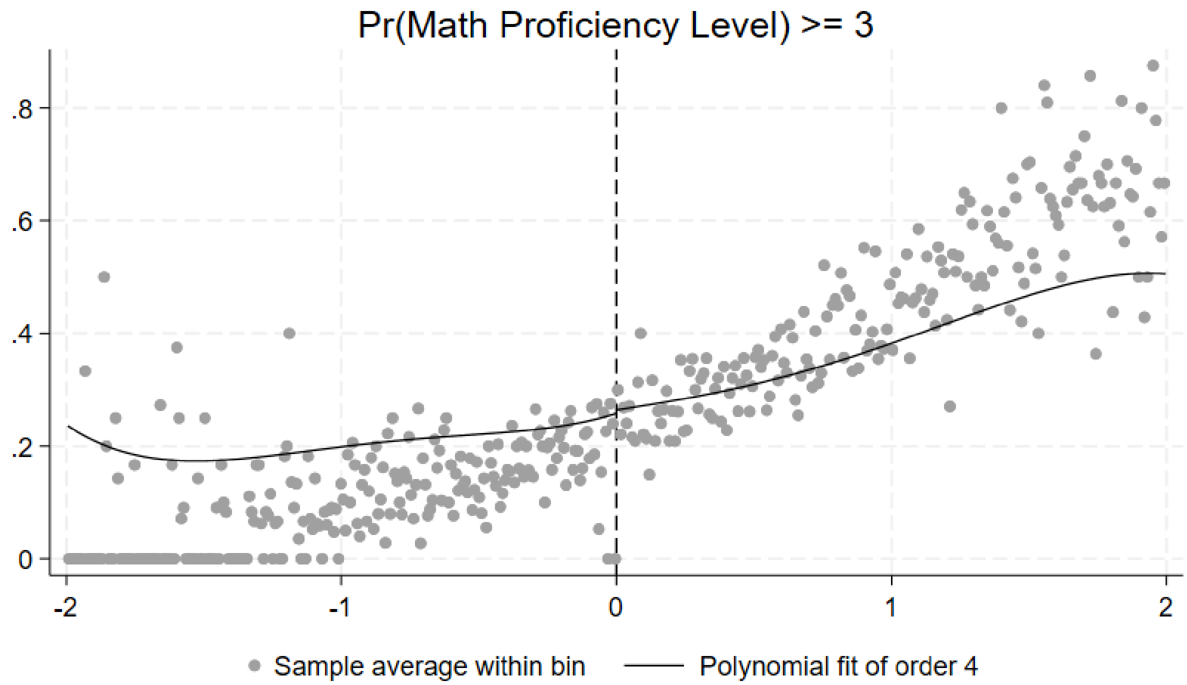
Note. Std. = standardized. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 6: Change in Probability of Scoring Proficient (3) or Above on SBAC ELA at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



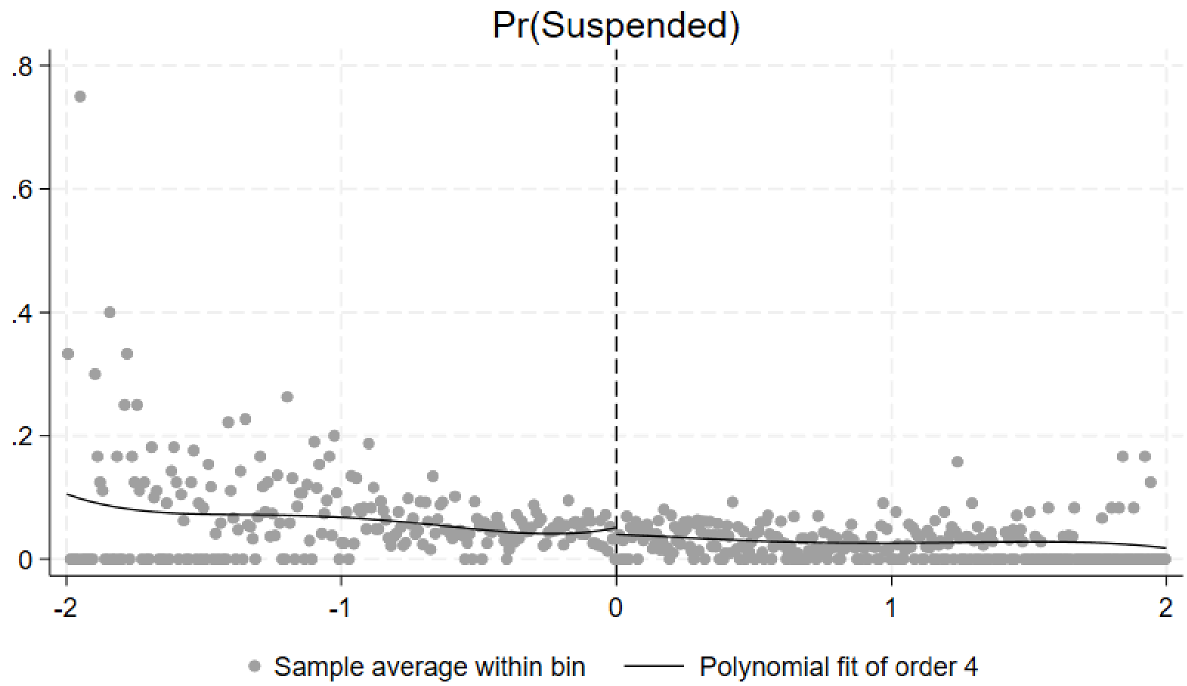
Note. SBAC = Smarter Balanced Assessment Consortium. Pr = probability. ELA = English language arts. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflects results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 7: Change in Probability of Scoring Proficient (3) or above on SBAC Math at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



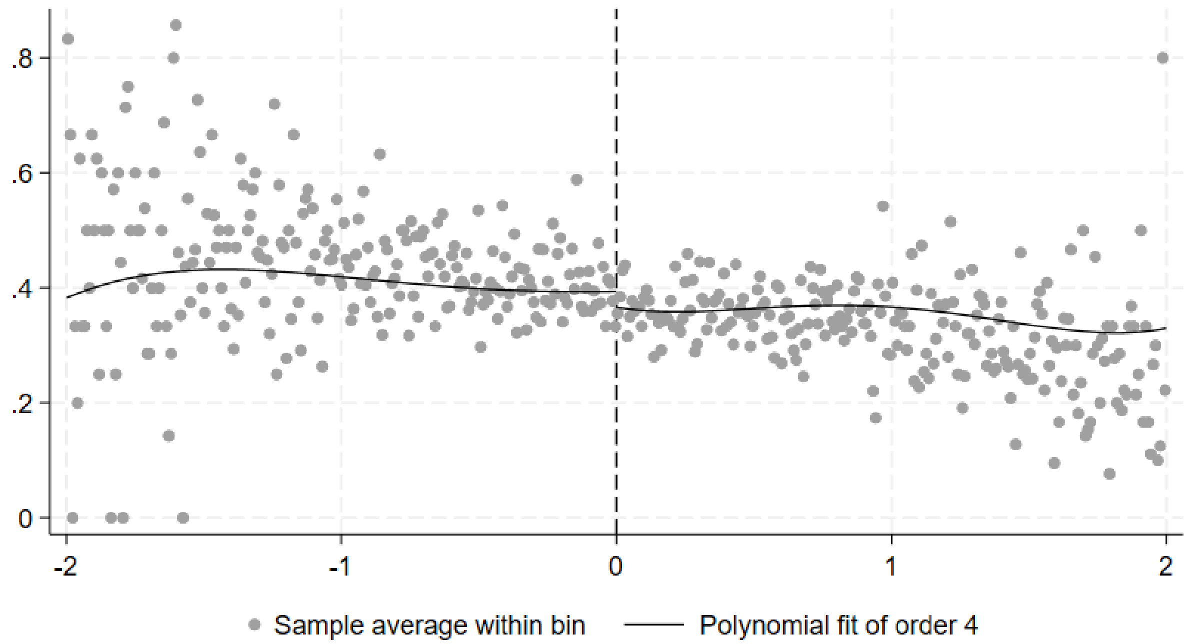
Note. SBAC = Smarter Balanced Assessment Consortium. Pr = probability. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 8: Change in Probability of Receiving a Suspension at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



Note. Pr = probability. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 9: Change in Probability of Being Chronically Absent (Missing More Than 10 Days) at the Margin of Basic Skills Criterion (Intent-to-Treat, Adjusted Model 4)



Note. Pr = probability. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Point estimates across outcomes were very small in magnitude in addition to not being statistically significant, indicating little evidence of an effect of retaining EL status among English proficient students at the margin of the academic criterion. There were two exceptions worth noting although in both cases the size of the standard errors relative to the point estimates indicate considerable uncertainty. First, point estimates indicate that being held in EL status one year out for this population may have a moderate, non-significant, positive impact on SBAC ELA scores of 0.107 SD. For a student at the mean, this would reflect an increase from the 50th to the 55th percentile in the distribution of SBAC ELA scores. The second exception was chronic absenteeism where the magnitude and direction of the estimate is consistent with the possibility that retaining EL status for this population may have had a large detrimental effect, increasing chronic absenteeism by 18.6%. The mean chronic absenteeism rate in our sample is 38%, so an 18.6 percentage point increase reflects a 50% increase to nearly 57%. The other outcomes examined: math scores, the likelihood of scoring proficient on both math and ELA SBAC tests, and suspension rates all had very small point estimates although they were all in the direction indicating negative effects of EL status retention. Our robustness checks, including adjusting for mass points (appendix Table A9), using uniform kernels (appendix Table A10), fitting local quadratic polynomials (appendix Table A11), using only robust standard errors (appendix Table A12), and testing placebo cutoffs (appendix Tables A13a and A13b), all aligned with these main results.

Prior research suggests that being held in EL status may have differential effects depending on grade level (Carlson & Knowles, 2016). As such, we divided our sample into elementary and secondary samples. These two samples are described in Table 8 and regression discontinuity results from our preferred model are shown in Table 9a (with estimates shown across bandwidths and model specifications standard errors presented in appendix Tables A13 and model specifications presented in appendix Tables A14 and A15). Here, too, we find all null results, suggesting no discernible effects of being held in EL status one year out for this population across outcomes and grade bands. Table 9b provides difference-in-regression discontinuity results confirming no statistically significant differences between the effects of being held in EL status between the elementary and secondary grade band samples.

Table 8: Student Characteristics in Grade-Band Level Sub-Samples

	Reclassified in grades 3-5	Reclassified in grades >=6
N	11,938 (50.4%)	11,755 (49.6%)
Entered Current-Year as Reclassified		
No	1,696 (14.2%)	2,988 (25.4%)
Yes	10,242 (85.8%)	8,767 (74.6%)
Current-Year Grade level*		
4	2,859 (23.9%)	
5	4,309 (36.1%)	
6	4,769 (39.9%)	
7		3,582 (30.5%)
8		4,201 (35.7%)
9		2,511 (21.4%)
11		10 (0.1%)
12		1,451 (12.3%)
Past-Year SBAC Tested Grade*		
3	2,860 (24.0%)	

	Reclassified in grades 3-5	Reclassified in grades >=6
4	4,308 (36.1%)	
5	4,767 (39.9%)	
6		3,582 (30.5%)
7		4,201 (35.7%)
8		2,511 (21.4%)
11		1,461 (12.4%)
Academic Year		
2022-2023	6,073 (50.9%)	5,720 (48.7%)
2023-2024	5,865 (49.1%)	6,035 (51.3%)
Gender		
Female	5,481 (45.9%)	5,345 (45.5%)
Male	6,457 (54.1%)	6,407 (54.5%)
Low Socioeconomic Status		
No	2,725 (22.8%)	2,027 (17.2%)
Yes	9,213 (77.2%)	9,728 (82.8%)
Race/Ethnicity		
Black	65 (0.5%)	64 (0.5%)
Asian	2,892 (24.2%)	1,988 (16.9%)
Filipino	144 (1.2%)	161 (1.4%)
Hispanic	7,966 (66.7%)	8,824 (75.1%)
Other	219 (1.8%)	241 (2.1%)
White	652 (5.5%)	477 (4.1%)
Language		
Arabic	195 (1.6%)	225 (1.9%)
Chinese	766 (6.4%)	395 (3.4%)
Other	2,096 (17.6%)	1,830 (15.6%)
Russian	138 (1.2%)	85 (0.7%)
Spanish	7,974 (66.8%)	8,790 (74.8%)
Vietnamese	769 (6.4%)	430 (3.7%)
Special Education		
No	11,317 (94.8%)	10,849 (92.3%)
Yes	621 (5.2%)	906 (7.7%)
Years in EL		
<=2	160 (1.3%)	96 (0.8%)
3--5	5,061 (42.4%)	709 (6.0%)
5+	6,717 (56.3%)	10,950 (93.2%)
Newcomer Status		
No	11,515 (96.5%)	11,485 (97.7%)
Yes	423 (3.5%)	270 (2.3%)
Past-Year Math Performance Level		
1	2,815 (23.6%)	6,710 (57.2%)
2	4,057 (34.0%)	3,377 (28.8%)
3	3,152 (26.4%)	1,089 (9.3%)

	Reclassified in grades 3-5	Reclassified in grades >=6
4	1,898 (15.9%)	531 (4.5%)
Missing	<10	22 (0.2%)
Student-Level Outcomes		
Std. ELA Score	0.223 (0.672)	-0.164 (0.651)
Std. Math Score	0.205 (0.763)	-0.194 (0.745)
Prop. ELA Proficiency Met or Exceeded	0.551 (0.497)	0.318 (0.466)
Prop. Math Proficiency Met or Exceeded	0.393 (0.489)	0.176 (0.381)
Prop. Chronically Absent	0.347 (0.476)	0.410 (0.492)
Prop. Any Suspension	0.022 (0.147)	0.059 (0.235)
District Characteristics		
Locale		
City	5,912 (49.5%)	5,895 (50.1%)
Suburb	4,489 (37.6%)	4,247 (36.1%)
Town	1,333 (11.2%)	1,414 (12.0%)
Rural	204 (1.7%)	199 (1.7%)
%Ever-EL	37.063 (14.910)	36.420 (14.519)
%Low-SES	64.374 (23.948)	65.699 (21.522)
%Latinx	56.6 (26)	57.9% (24)
District-Size	21,792.558 (18,428.332)	21,946.169 (17,774.954)
District Avg. ELA Proficient	0.471 (0.16)	0.47 (0.14)
District Avg. Math Proficient	0.364 (0.19)	0.341 (0.16)
District Avg. Prop. Any Suspension	0.035 (0.02)	0.037 (0.02)
District Avg. Prop. Chronically Absent	0.461 (0.1)	0.47 (0.1)

Note. ELPAC = English language proficiency assessment of California. EL = English learner. ELA = English language arts. SES = Socioeconomic status. Table reports frequencies (percentage) for categorical variables and mean (standard deviation) for continuous variables.

*We restricted our sample to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past-year.

Table 9A: Estimates of the Effect of EL-status Retention among Students who Met District ELPAC criteria for Students on the Margin of Basic Skills Criteria across Outcomes, Optimal Bandwidths, Preferred Model Specification, by Grade Band

	ITT				TOT			
Outcome	Optima l BW	N in Optimal BW	Estimate	Std. Error	Optimal BW	N in Optima l BW	Estimate	Std. Error
Elementary Sample: Reclassified in grades 3-5								
Std. ELA Score	0.55	5329	0.025	0.041	0.76	7029	0.079	0.259
Std. Math Score	0.657	6213	0.012	0.051	0.778	7141	0.005	0.331
Pr(ELA >= 3)	0.595	5719	0.019	0.029	0.636	6028	0.118	0.198
Pr(Math >= 3)	0.704	6591	-0.004	0.029	0.786	7184	-0.047	0.196
Pr(Suspension)	0.617	5980	0.000	0.009	0.581	5676	0.001	0.059
Pr(Chronic Absenteeism)	0.611	5921	0.022	0.028	0.822	7544	0.197	0.201
Secondary Sample: Reclassified in grades >= 6								
Std. ELA Score	0.745	4851	0.012	0.058	0.66	4427	0.089	0.289
Std. Math Score	0.693	4592	-0.028	0.063	0.807	5147	-0.132	0.315
Pr(ELA >= 3)	0.803	5141	0.000	0.042	0.653	4382	-0.012	0.222
Pr(Math >= 3)	0.659	4404	-0.014	0.029	0.631	4248	-0.073	0.156
Pr(Suspension)	0.762	7505	0.007	0.011	0.728	7219	0.038	0.059
Pr(Chronic Absenteeism)	0.589	6102	0.037	0.029	0.628	6452	0.189	0.143

Note. ELPAC = English language proficiency assessment for California. FE = fixed effects. ITT = intent-to-treat. TOT = treatment-on-the-treated. Std. = standardized. ELA = English language arts. PL = proficiency level. SD = standard deviation. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model). Models estimated using rdrobust package in Stata with triangular kernels and linear fits. Estimates show the change in average outcome at the margin of the SBAC threshold. Inference based on heterosckedastic-robust and clustered standard errors. Covariates include flags for race/ethnicity, language code, indicator for low socio-economic status, indicator for special education placement, grade, newcomer status, and past year ELP scaled score. Estimates across other bandwidths and model specification are presented in the appendix. p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

Table 9B: Summary of the Difference in Regression Discontinuity Estimates between Elementary and Secondary Grade Samples across Outcomes, Optimal Bandwidth, and Preferred Model Specification

	ITT						TOT					
	Secondary		Elementary		Difference		Secondary		Elementary		Difference	
Outcome	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Secondary - Elementary												
Std. ELA Score	0.012	0.058	0.025	0.041	-0.013	0.071	0.089	0.289	0.079	0.259	0.01	0.388
Std. Math Score	-0.028	0.063	0.012	0.051	-0.04	0.081	-0.132	0.315	0.005	0.331	-0.137	0.457
Pr(ELA >= 3)	0.000	0.042	0.019	0.029	-0.019	0.051	-0.012	0.222	0.118	0.198	-0.13	0.297
Pr(Math >= 3)	-0.014	0.029	-0.004	0.029	-0.01	0.041	-0.073	0.156	-0.047	0.196	-0.026	0.251
Pr(Suspension)	0.007	0.011	0.000	0.009	0.007	0.014	0.038	0.059	0.001	0.059	0.037	0.083
Pr(Chronic Absenteeism)	0.037	0.029	0.022	0.028	0.015	0.040	0.189	0.143	0.197	0.201	-0.008	0.247

Note. ITT = intent-to-treat. TOT = treatment-on-the-treated. SE = standard error. Std = standardized. ELA = English language arts. Pr = probability. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model).
p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

While non-significant, we briefly note results of meaningful magnitude, keeping in mind these are highly imprecise and therefore should be interpreted cautiously. Following Kraft's (2020) guidelines on educational effect sizes, the point estimates are consistent with potentially positive effects of retaining EL status on ELA scores (0.079 SD) and ELA proficiency likelihood (11.8 percentage points) in elementary grades and evidence of moderately sized negative effects of being held in EL status on math scores (-0.132 SD) and math proficiency likelihood (-7.3 percentage points) at the secondary level. Chronic absenteeism remains negative and of meaningful size in both the elementary and secondary samples. Although these estimates are imprecise and not statistically significant, their magnitude suggests that future research may be warranted to examine whether retaining EL status positively influences ELA outcomes in the early grades, negatively influences math outcomes in the secondary grades, and increases chronic absenteeism across grade bands.

Research Question 4: What, if any, are the Differences in the Effects of Being Held in EL Status at the Margin of the Basic Skills Criterion for Students in Districts with Lenient Versus Strenuous Basic Skills Thresholds?

Because California allows for local control of the basic skills reclassification criterion, we can assess whether reclassification effects differ systematically by difficulty level. Following conventional frameworks assessing the appropriateness of reclassification thresholds, one might expect negative effects of reclassification if the academic threshold is set too low, positive effects if set too high, and null effects if appropriately set (see Figure 1). California's tiered thresholds – with some districts setting the academic bar far lower than others – is an optimal setting to assess the relevancy of this framework. Importantly, we find clear evidence of differences in reclassification effects by district basic skills criterion type. However, these results do not support the conventional framework for understanding reclassification effects. Instead, they point to a more complex relationship of reclassification criteria to student outcomes, which we will describe more in the discussion section.

As described above, we assess heterogeneity in reclassification effects at the margin of the basic skills threshold in two ways. First, we use the “SBAC Or” “SBAC Only” and “SBAC And” categories, understanding that these reflect increasing levels of difficulty. Second, we divide districts by SBAC proficiency level threshold, focusing on SBAC proficiency levels 2 and 3, against increasing difficulty. Tables 10 and 12 show descriptive statistics of these subgroups, respectively, while Tables 11a and 13a present regression discontinuity results from our preferred model specification (with estimates and standard errors from models across bandwidth and specification presented in appendix Tables A16 through A19). In addition to these main tables of regression discontinuity results, Tables 11b and 13b show our difference-in-regression discontinuity results by difficulty level and threshold, respectively.

Table 10: Student Characteristics across SBAC-Difficulty Sub-Samples

	SBAC-Or	Only SBAC	SBAC-And
N	19,482 (82.2%)	2,522 (10.6%)	1,689 (7.1%)
Entered Current-Year as Reclassified			
No	3,259 (16.7%)	816 (32.4%)	609 (36.1%)
Yes	16,223 (83.3%)	1,706 (67.6%)	1,080 (63.9%)
Current-Year Grade level*			
4	2,386 (12.2%)	231 (9.2%)	242 (14.3%)
5	3,648 (18.7%)	338 (13.4%)	323 (19.1%)
6	4,076 (20.9%)	400 (15.9%)	293 (17.3%)
7	2,934 (15.1%)	364 (14.4%)	284 (16.8%)
8	3,459 (17.8%)	436 (17.3%)	306 (18.1%)
9	1,954 (10.0%)	399 (15.8%)	158 (9.4%)
12	1,014 (5.2%)	354 (14.0%)	83 (4.9%)
Past-Year SBAC Tested Grade*			
3	2,387 (12.3%)	231 (9.2%)	242 (14.3%)
4	3,647 (18.7%)	338 (13.4%)	323 (19.1%)
5	4,074 (20.9%)	400 (15.9%)	293 (17.3%)
6	2,937 (15.1%)	364 (14.4%)	284 (16.8%)
7	3,459 (17.8%)	436 (17.3%)	306 (18.1%)
8	1,954 (10.0%)	399 (15.8%)	158 (9.4%)
11	1,024 (5.3%)	354 (14.0%)	83 (4.9%)
Academic Year			
2022-2023	9,821 (50.4%)	1,187 (47.1%)	785 (46.5%)
2023-2024	9,661 (49.6%)	1,335 (52.9%)	904 (53.5%)
Gender			
Female	8,905 (45.7%)	1,165 (46.2%)	756 (44.8%)

	SBAC-Or	Only SBAC	SBAC-And
Male	10,574 (54.3%)	1,357 (53.8%)	933 (55.2%)
Low Socioeconomic Status			
No	3,464 (17.8%)	667 (26.4%)	621 (36.8%)
Yes	16,018 (82.2%)	1,855 (73.6%)	1,068 (63.2%)
Race/Ethnicity			
African American	94 (0.5%)	30 (1.2%)	<10
Asian	3,669 (18.8%)	547 (21.7%)	664 (39.3%)
Filipino	255 (1.3%)	34 (1.3%)	16 (0.9%)
Hispanic	14,304 (73.4%)	1,689 (67.0%)	797 (47.2%)
Other	367 (1.9%)	62 (2.5%)	31 (1.8%)
White	793 (4.1%)	160 (6.3%)	176 (10.4%)
Language			
Arabic	289 (1.5%)	101 (4.0%)	30 (1.8%)
Chinese	934 (4.8%)	52 (2.1%)	175 (10.4%)
Other	2,671 (13.7%)	646 (25.6%)	609 (36.1%)
Russian	160 (0.8%)	24 (1.0%)	39 (2.3%)
Spanish	14,303 (73.4%)	1,659 (65.8%)	802 (47.5%)
Vietnamese	1,125 (5.8%)	40 (1.6%)	34 (2.0%)
Special Education			
No	18,234 (93.6%)	2,368 (93.9%)	1,564 (92.6%)
Yes	1,248 (6.4%)	154 (6.1%)	125 (7.4%)
Years in EL			
<=2	168 (0.9%)	31 (1.2%)	57 (3.4%)
3--5	4,791 (24.6%)	516 (20.5%)	463 (27.4%)
5+	14,523 (74.5%)	1,975 (78.3%)	1,169 (69.2%)
Newcomer Status			
No	19,026 (97.7%)	2,432 (96.4%)	1,542 (91.3%)
Yes	456 (2.3%)	90 (3.6%)	147 (8.7%)
Past-Year Math Performance Level			
1	8,043 (41.3%)	961 (38.1%)	521 (30.9%)
2	6,274 (32.3%)	701 (27.8%)	459 (27.2%)
3	3,386 (17.4%)	498 (19.8%)	357 (21.2%)
4	1,725 (8.9%)	359 (14.2%)	345 (20.5%)
Missing	25 (0.1%)	<10	<10
Student-Level Outcomes			
Std. ELA Score	0.044 (0.686)	0.190 (0.668)	0.236 (0.728)
Std. Math Score	-0.006 (0.760)	0.283 (0.769)	0.396 (0.894)
Prop. ELA Proficiency Met or Exceeded	0.442 (0.497)	0.531 (0.499)	0.571 (0.495)
Prop. Math Proficiency Met or Exceeded	0.281 (0.449)	0.417 (0.493)	0.491 (0.500)

	SBAC-Or	Only SBAC	SBAC-And
Prop. Chronically Absent	0.385 (0.487)	0.370 (0.483)	0.321 (0.467)
Prop. Any Suspension	0.043 (0.203)	0.023 (0.151)	0.035 (0.184)
District Characteristics			
Locale			
City	9,339 (47.9%)	1,307 (51.8%)	1,161 (68.7%)
Suburb	7,275 (37.3%)	948 (37.6%)	513 (30.4%)
Town	2,623 (13.5%)	109 (4.3%)	15 (0.9%)
Rural	245 (1.3%)	158 (6.3%)	<10
%Ever-EL	38.076 (15.143)	32.807 (12.116)	27.252 (6.464)
%Low-SES	67.789 (20.710)	57.575 (25.567)	44.357 (27.612)
%Latinx	59.7 (24.3)	54.3 (23)	33.6 (23.8)
District-Size	23,433.142 (19,438.897)	14,599.247 (7,033.103)	14,679.105 (2,816.323)
District Avg. ELA Proficient	0.45 (0.133)	0.532 (0.141)	0.618 (0.192)
District Avg. Math Proficient	0.328 (0.15)	0.417 (0.178)	0.543 (0.263)
District Avg. Prop. Any Suspension	0.039 (0.017)	0.024 (0.011)	0.026 (0.019)
District Avg. Prop. Chronically Absent	0.476 (0.094)	0.436 (0.105)	0.388 (0.118)

Note. EL = English learner. ELA = English language arts. SES = socioeconomic status. Prop. = proportion. SBAC = Smarter Balanced Assessment Consortium. Ever-EL refers to all current and former (reclassified) EL students. Low-SES refers to students eligible for free- or reduced-price lunch. Table reports frequencies (percentage) for categorical variables and mean (standard deviation) for continuous variables. “Or” districts are districts that require that students meet a given threshold on the named assessment (ELPAC for English language proficiency [ELP] criterion, SBAC for basic skills criterion) or a threshold on one or more alternative assessments. “Only” districts require only the named assessment. “And” districts require that students meet the threshold on the named assessment and meet one or more additional thresholds on other assessments or assessment components.

*We restricted our sample to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past-year.

Table 11A: Estimates of the Effect of EL-status Retention among Students who Met District ELPAC criteria for Students on the Margin of Basic Skills Criteria across Outcomes, Optimal Bandwidths and Preferred Model Specification, by District Basic Skills Difficulty

	ITT				TOT			
Outcome	Optimal BW	N in Optimal BW	Estimate	Std. Error	Optimal BW	N in Optimal BW	Estimate	Std. Error
Academic Criteria: SBAC Or								
Std. ELA Score	0.59	5665	0.026	0.057	1.088	9103	-0.008	0.323
Std. Math Score	0.843	7627	0.027	0.06	1.008	8639	0.133	0.331
Pr(ELA >= 3)	0.837	7579	0.015	0.041	0.987	8520	0.05	0.229
Pr(Math >= 3)	0.742	6851	0.01	0.034	0.971	8420	0.04	0.182
Pr(Suspension)	0.775	8544	0.004	0.009	0.583	6758	0.033	0.056
Pr(Chronic Absenteeism)	0.626	7158	0.009	0.03	0.682	7699	0.067	0.161
Academic Criteria: SBAC Only								
Std. ELA Score	0.56	606	0.240**	0.092	0.545	585	0.485**	0.147
Std. Math Score	0.651	684	-0.113	0.094	0.502	544	-0.187	0.258
Pr(ELA >= 3)	0.72	758	0.071	0.071	0.555	598	-0.043	0.138
Pr(Math >= 3)	0.421	469	0.046	0.09	0.756	783	-0.044	0.144
Pr(Suspension)	0.794	1102	-0.009	0.012	0.61	896	-0.012	0.032
Pr(Chronic Absenteeism)	0.538	806	0.039	0.072	0.553	826	0.079	0.144
Academic Criteria: SBAC And								
Std. ELA Score	0.937	990	0.058	0.108	0.618	703	0.393	0.354
Std. Math Score	0.393	463	0.148*	0.069	0.509	589	0.427~	0.244
Pr(ELA >= 3)	0.666	755	-0.022	0.079	0.611	693	-0.065	0.287
Pr(Math >= 3)	0.56	644	0.004	0.082	0.408	478	0.124	0.284
Pr(Suspension)	0.908	1167	-0.003	0.021	0.573	815	-0.052	0.132
Pr(Chronic Absenteeism)	0.718	969	0.023	0.041	0.622	868	0.101	0.189

Note. SBAC = Smarter Balanced Assessment Consortium. FE = fixed effects. ITT = intent-to-treat. TOT = treatment-on-the-treated. Std. = standardized. ELA = English language arts. PL = proficiency level. SD = standard deviation. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model). “Or” districts are districts that require that students meet a given threshold on the named assessment (ELPAC for English language proficiency [ELP] criterion, SBAC for basic skills criterion) or a threshold on one or more alternative assessments. “Only” districts require only the named assessment. “And” districts require that students meet the threshold on the named assessment and meet one or more additional thresholds on other assessments or assessment components. Models estimated using *rdrobust* package in Stata with triangular kernels and linear fits. Estimates show the change in average outcome at the margin of the SBAC threshold. Inference based on heterosckedastic-robust and clustered standard errors. Covariates include flags for race/ethnicity, language code, indicators for low socio-economic status, indicators for special education placement, grade, newcomer status, and past year ELP scaled score.
p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

Table 11B. Summary of the Difference in Regression Discontinuity Estimates by District SBAC Difficulty across Outcomes, Optimal Bandwidth, and Preferred Model Specification

	ITT						TOT					
	Subgroup 1		Subgroup 2		Difference		Subgroup 1		Subgroup 2		Difference	
Outcome	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Academic Criteria: SBAC Only (Subgroup 1) - SBAC Or (Subgroup 2)												
Std. ELA Score	0.240**	0.092	0.026	0.057	0.214*	0.108	0.485**	0.147	-0.008	0.323	0.493	0.355
Std. Math Score	-0.113	0.094	0.027	0.06	-0.14	0.112	-0.187	0.258	0.133	0.331	-0.32	0.420
Pr(ELA >= 3)	0.071	0.071	0.015	0.041	0.056	0.082	-0.043	0.138	0.05	0.229	-0.093	0.267
Pr(Math >= 3)	0.046	0.09	0.01	0.034	0.036	0.096	-0.044	0.144	0.04	0.182	-0.084	0.232
Pr(Suspension)	-0.009	0.012	0.004	0.009	-0.013	0.015	-0.012	0.032	0.033	0.056	-0.045	0.064
Pr(Chronic Absenteeism)	0.039	0.072	0.009	0.03	0.03	0.078	0.079	0.144	0.067	0.161	0.012	0.216
Academic Criteria: SBAC And (Subgroup 1) - SBAC Or (Subgroup 2)												
Std. ELA Score	0.058	0.108	0.026	0.057	0.032	0.122	0.393	0.354	-0.008	0.323	0.401	0.479
Std. Math Score	0.148*	0.069	0.027	0.06	0.121	0.091	0.427~	0.244	0.133	0.331	0.294	0.411
Pr(ELA >= 3)	-0.022	0.079	0.015	0.041	-0.037	0.089	-0.065	0.287	0.05	0.229	-0.115	0.367
Pr(Math >= 3)	0.004	0.082	0.01	0.034	-0.006	0.089	0.124	0.284	0.04	0.182	0.084	0.337
Pr(Suspension)	-0.003	0.021	0.004	0.009	-0.007	0.023	-0.052	0.132	0.033	0.056	-0.085	0.143
Pr(Chronic Absenteeism)	0.023	0.041	0.009	0.03	0.014	0.051	0.101	0.189	0.067	0.161	0.034	0.248
Academic Criteria: SBAC And (Subgroup 1) - SBAC Only (Subgroup 2)												
Std. ELA Score	0.058	0.108	0.240**	0.092	-0.182	0.142	0.393	0.354	0.485**	0.147	-0.092	0.383
Std. Math Score	0.148*	0.069	-0.113	0.094	0.261	0.117	0.427~	0.244	-0.187	0.258	0.614	0.355
Pr(ELA >= 3)	-0.022	0.079	0.071	0.071	-0.093	0.106	-0.065	0.287	-0.043	0.138	-0.022	0.318
Pr(Math >= 3)	0.004	0.082	0.046	0.09	-0.042	0.122	0.124	0.284	-0.044	0.144	0.168	0.318
Pr(Suspension)	-0.003	0.021	-0.009	0.012	0.006	0.024	-0.052	0.132	-0.012	0.032	-0.04	0.136
Pr(Chronic Absenteeism)	0.023	0.041	0.039	0.072	-0.016	0.083	0.101	0.189	0.079	0.144	0.022	0.238

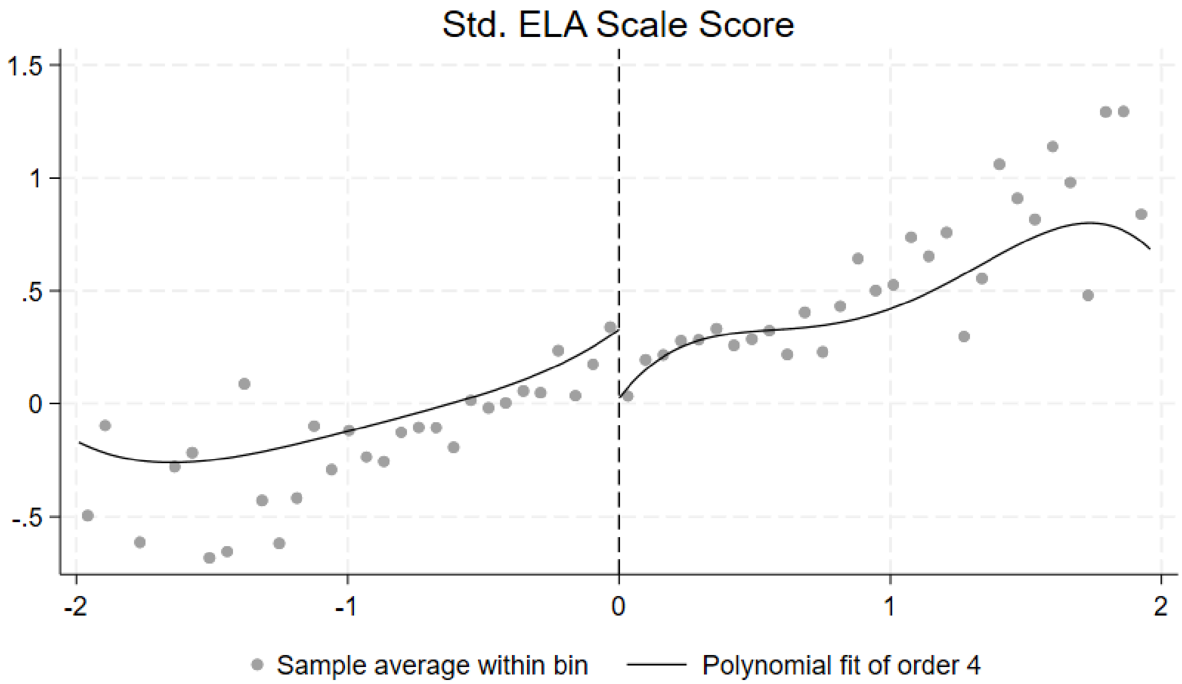
Note. ITT = intent-to-treat. TOT = treatment-on-the-treated. Est = estimate. SE = standard error. SBAC = Smarter Balanced Assessment Consortium. Std = standardized. ELA = English language arts. Pr = probability. . "Or" districts are districts that require that students meet a given threshold on the named assessment (ELPAC for English language proficiency [ELP] criterion, SBAC for basic skills criterion) or a threshold on one or more alternative assessments. "Only" districts require only the named assessment. "And" districts require that students meet the threshold on the named assessment and meet one or more additional thresholds on other assessments or assessment components. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model).

p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

EL Status Retention Effects by District Basic Skills Criterion Difficulty

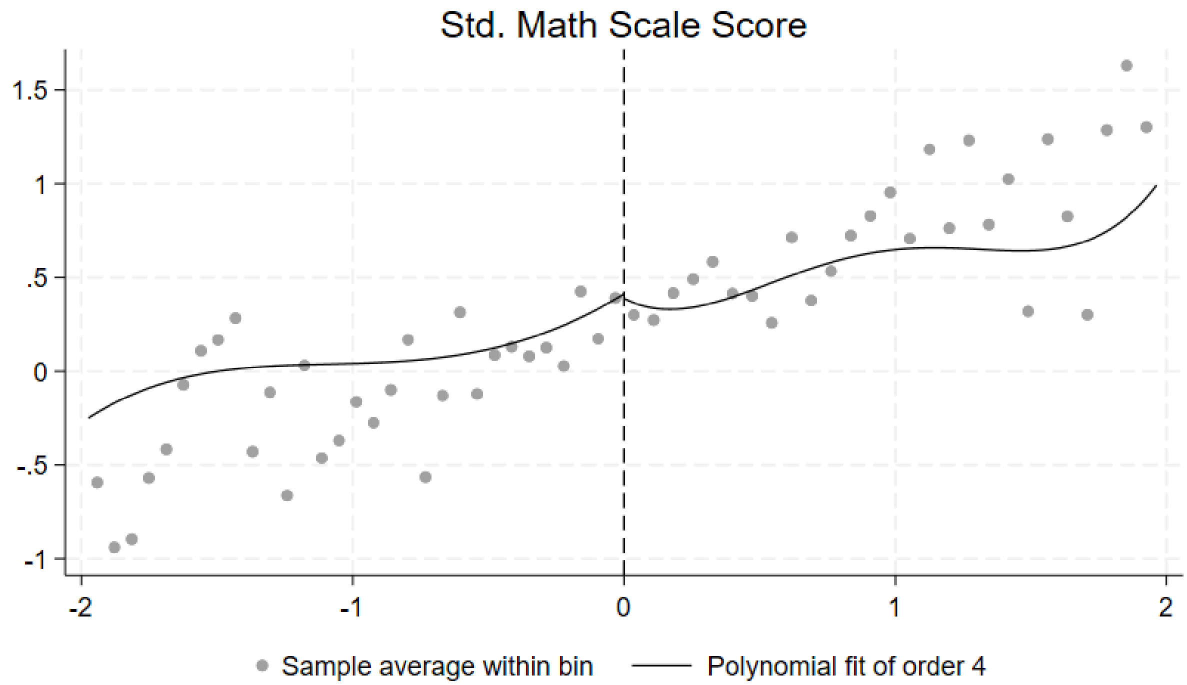
Our results indicate that reclassification effects differ by basic skills criterion difficulty, particularly regarding subsequent year SBAC test score results. Specifically, we find that there are no measurable effects of being held in EL status in districts that have more lenient basic skills thresholds ("SBAC or" districts) but that there are positive effects of retaining EL status for students in districts with more strenuous criteria ("SBAC only" and "SBAC and" districts). As a reminder, "SBAC or" districts make up most districts. In "SBAC or" districts we observe non-significant and, in general, very small point estimates. However, in those districts with more strenuous basic skills criteria we find sizable positive effects of being held in EL status on subsequent year ELA score (0.485 SD TOT estimate in "SBAC only" districts; significant at .05 level), and math (0.427 SD TOT estimate in "SBAC and" districts; marginally significant). Difference-in-regression discontinuity results confirm that these differences are statistically significant in the case of ELA between "SBAC only" and "SBAC or" districts (see Table 11b). Results from unadjusted ITT models are shown visually in figures 10 and 11 for "SBAC only" and "SBAC and", respectively (parallel figures for the other subgroups are included in appendix, Figures A1 through A4). Assuming normal distribution of SBAC scores and that point estimates just below and above the cutoff are close to the mean, this close to half of a standard deviation boost is equivalent to going from the 50th to the 68th percentile (SBAC, 2026).

Figure 10: Change in Standardized ELA Score at the Margin of Basic Skills Criteria among "SBAC Only" Districts (Intent-to-Treat, Adjusted Model 4)



Note. SBAC = Smarter Balanced Assessment Consortium. ELA = English language arts. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

Figure 11: Change in Standardized Math Score at the Margin of Basic Skills Criteria among "SBAC And" Districts (Intent-to-Treat, Adjusted Model 4)



Note. SBAC = Smarter Balanced Assessment Consortium. Std. = standardized. The figure is adjusted for student characteristics and district fixed effect using *rdrobust* Stata package and reflect results shown in ITT Model 4. Regression lines do not visually align perfectly with scatterplot because lines reflect adjusted results whereas scatterplot depicts raw data.

EL Status Retention Effects by SBAC Threshold

Results from analyses by SBAC proficiency level threshold mirror those found by criterion operationalization. Namely, we continue to find no statistically significant effects of reclassification on subsequent outcomes among English proficient students at the margin of the ELA threshold with the exception of subsequent academic (ELA) test scores for students in districts with more strenuous thresholds. Specifically, in districts that set their SBAC proficiency level threshold to 3 (considerably above the state mean), students retained in EL status experience a positive (0.305 SD) boost on their subsequent year SBAC ELA test. We do not see parallel effects in districts with more lenient thresholds (SBAC level 2). The differences across threshold subgroups, however, do not reach statistical significance in our difference-in-regression discontinuity analyses (see Table 13b). The differences are, however, meaningful in magnitude. Assuming normal distribution of SBAC scores and that point estimates just below and above the cutoff are close to the mean, a 0.31 standard deviation boost is equivalent to going from the 50th to the 60th percentile (SBAC, 2026). It bears noting that our chronic absenteeism estimates, while never statistically significant, were negative across all models and basic skills criteria difficulty levels.

Table 12: Student Characteristics across SBAC-Threshold Sub-Samples

	SBAC Thresh 2	Thresh 3	Between 2 and 3
Student Characteristics			
N	9,196 (38.8%)	8,304 (35.0%)	6,193 (26.1%)
Entered Current-Year as Reclassified			
No	1,314 (14.3%)	2,074 (25.0%)	1,296 (20.9%)
Yes	7,882 (85.7%)	6,230 (75.0%)	4,897 (79.1%)
Current-Year Grade level*			
4	1,151 (12.5%)	1,052 (12.7%)	656 (10.6%)
5	1,681 (18.3%)	1,511 (18.2%)	1,117 (18.0%)
6	1,862 (20.2%)	1,703 (20.5%)	1,204 (19.4%)
7	1,351 (14.7%)	1,309 (15.8%)	922 (14.9%)
8	1,610 (17.5%)	1,445 (17.4%)	1,146 (18.5%)
9	997 (10.8%)	791 (9.5%)	723 (11.7%)
12	541 (5.9%)	486 (5.9%)	424 (6.8%)
Past-Year SBAC Tested Grade*			
3	1,152 (12.5%)	1,052 (12.7%)	656 (10.6%)
4	1,681 (18.3%)	1,510 (18.2%)	1,117 (18.0%)
5	1,861 (20.2%)	1,703 (20.5%)	1,203 (19.4%)
6	1,352 (14.7%)	1,310 (15.8%)	923 (14.9%)
7	1,610 (17.5%)	1,445 (17.4%)	1,146 (18.5%)
8	997 (10.8%)	791 (9.5%)	723 (11.7%)
11	543 (5.9%)	493 (5.9%)	425 (6.9%)
Academic Year			
2022-2023	4,408 (47.9%)	4,299 (51.8%)	3,086 (49.8%)
2023-2024	4,788 (52.1%)	4,005 (48.2%)	3,107 (50.2%)
Gender			
Female	2,879 (46.5%)	4,215 (45.8%)	3,732 (44.9%)
Male	3,313 (53.5%)	4,981 (54.2%)	4,570 (55.0%)
Low Socioeconomic Status			
No	1,755 (19.1%)	2,634 (31.7%)	363 (5.9%)
Yes	7,441 (80.9%)	5,670 (68.3%)	5,830 (94.1%)
Race/Ethnicity			
African American	62 (0.7%)	49 (0.6%)	18 (0.3%)
Asian	1,800 (19.6%)	2,650 (31.9%)	430 (6.9%)

	SBAC Thresh 2	Thresh 3	Between 2 and 3
Filipino	176 (1.9%)	105 (1.3%)	24 (0.4%)
Hispanic	6,519 (70.9%)	4,717 (56.8%)	5,554 (89.7%)
Other	243 (2.6%)	169 (2.0%)	48 (0.8%)
White	396 (4.3%)	614 (7.4%)	119 (1.9%)
Language			
Arabic	97 (1.1%)	221 (2.7%)	102 (1.6%)
Chinese	435 (4.7%)	579 (7.0%)	147 (2.4%)
Other	1,765 (19.2%)	1,848 (22.3%)	313 (5.1%)
Russian	71 (0.8%)	147 (1.8%)	<10
Spanish	6,506 (70.7%)	4,673 (56.3%)	5,585 (90.2%)
Vietnamese	322 (3.5%)	836 (10.1%)	41 (0.7%)
Special Education			
No	8,629 (93.8%)	7,668 (92.3%)	5,869 (94.8%)
Yes	567 (6.2%)	636 (7.7%)	324 (5.2%)
Years in EL			
<=2	83 (0.9%)	145 (1.7%)	28 (0.5%)
3-5	2,262 (24.6%)	2,152 (25.9%)	1,356 (21.9%)
5+	6,851 (74.5%)	6,007 (72.3%)	4,809 (77.7%)
Newcomer Status			
No	8,979 (97.6%)	7,904 (95.2%)	6,117 (98.8%)
Yes	217 (2.4%)	400 (4.8%)	76 (1.2%)
Past-Year Math Performance Level			
1	3,898 (42.5%)	2,700 (32.6%)	2,927 (47.3%)
2	2,997 (32.6%)	2,448 (29.5%)	1,989 (32.1%)
3	1,583 (17.2%)	1,749 (21.1%)	909 (14.7%)
4	685 (7.5%)	1,386 (16.7%)	358 (5.8%)
Missing	18 (0.2%)	<10	<10
Student-Level Outcomes			
Std. ELA Score	0.010 (0.676)	0.213 (0.705)	-0.034 (0.659)
Std. Math Score	-0.039 (0.733)	0.276 (0.833)	-0.133 (0.700)
Prop. ELA Proficiency Met or Exceeded	0.419 (0.493)	0.549 (0.498)	0.395 (0.489)
Prop. Math Proficiency Met or Exceeded	0.265 (0.441)	0.419 (0.493)	0.220 (0.414)
Prop. Chronically Absent	0.404 (0.491)	0.343 (0.475)	0.389 (0.488)
Prop. Any Suspension	0.046 (0.209)	0.033 (0.178)	0.042 (0.201)
District Characteristics			
Locale			
City	3,379 (36.7%)	5,232 (63.0%)	3,196 (51.6%)
Suburb	4,814 (52.3%)	2,364 (28.5%)	1,558 (25.2%)
Town	979 (10.6%)	345 (4.2%)	1,423 (23.0%)
Rural	24 (0.3%)	363 (4.4%)	16 (0.3%)
%Ever-EL	38.271 (15.138)	33.336 (14.705)	39.045 (13.210)
%Low-SES	65.390 (19.348)	51.205 (23.250)	83.038 (11.179)

	SBAC Thresh 2	Thresh 3	Between 2 and 3
%Latinx	56.5 (24.8)	45.9 (23.5)	73.6 (17.6)
District-Size	26,835.115 (21,666.518)	14,037.558 (8,469.373)	24,994.861 (18,235.239)
District Avg. ELA Proficient	0.424 (0.117)	0.572 (0.148)	0.404 (0.107)
District Avg. Math Proficient	0.3 (0.129)	0.481 (0.187)	0.26 (0.101)
District Avg. Prop. Any Suspension	0.044 (0.015)	0.027 (0.017)	0.037 (0.013)
District Avg. Prop. Chronically Absent	0.5 (0.075)	0.416 (0.116)	0.48 (0.081)

Note. EL = English learner. ELA = English language arts. SES = socioeconomic status. Prop. = proportion. SBAC = Smarter Balanced Assessment Consortium. Ever-EL refers to all current and former (reclassified) EL students. Low-SES refers to students eligible for free- or reduced-price lunch. Table reports frequencies (percentage) for categorical variables and mean (standard deviation) for continuous variables.

*We restricted our sample to observations who had valid scores on SBAC from last year. Since SBAC is administered in Grades 3-8, and 11, most students would be in Grades 4-9, and 12, in the current year. We preserve observations where students repeated grades, therefore, a handful of observations in Grades 3 and 11 have valid SBAC scores from past-year.

Table 13A: Estimates of the Effect of EL-status Retention among Students who Met District ELPAC criteria for Students on the Margin of Basic Skills Criteria across Outcomes, Optimal Bandwidths, and Preferred Model Specification, by District SBAC Proficiency Level Threshold

	ITT				TOT			
Outcome	Optimal BW	N in Optimal BW	Estimate	Std. Error	Optimal BW	N in Optimal BW	Estimate	Std. Error
SBAC Threshold 2								
Std. ELA Score	0.461	2635	0.02	0.036	0.531	2994	0.118	0.188
Std. Math Score	0.679	3751	0.013	0.067	0.652	3628	0.073	0.346
Pr(ELA >= 3)	0.666	3690	-0.009	0.032	0.602	3374	-0.034	0.163
Pr(Math >= 3)	0.604	3379	0.014	0.028	0.547	3089	0.075	0.153
Pr(Suspension)	0.481	3428	-0.008	0.02	0.501	3567	-0.037	0.089
Pr(Chronic Absenteeism)	0.5	3558	0.005	0.038	0.61	4285	0.047	0.165
SBAC Threshold 3								
Std. ELA Score	0.466	2950	0.095~	0.051	0.663	4035	0.305	0.192
Std. Math Score	0.704	4220	0.027	0.058	0.629	3855	0.139	0.265
Pr(ELA >= 3)	0.567	3523	0.055	0.042	0.634	3878	0.235	0.18
Pr(Math >= 3)	0.652	3964	0.003	0.037	0.718	4267	0.003	0.17
Pr(Suspension)	0.642	4668	0.013	0.009	0.935	6095	0.054	0.042
Pr(Chronic Absenteeism)	0.601	4428	0.014	0.03	0.719	5090	0.082	0.127

Note. ELPAC = English language proficiency assessment for California. FE = fixed effects. ITT = intent-to-treat. TOT = treatment-on-the-treated. SBAC = Smarter Balanced Assessment Consortium. Std. = standardized. ELA = English language arts. PL = proficiency level. SD = standard deviation. Preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model). Models estimated using *rdrobust* package in Stata with triangular kernels and linear fits. Estimates show the change in average outcome at the margin of the SBAC threshold. Inference based on heterosckedastic-robust and clustered standard errors. Covariates include flags for race/ethnicity, language code, indicators for low socio-economic status, indicators for special education placement, grade, newcomer status, and past year ELP scaled score.
p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

Table 13B: Summary of the Difference in Regression Discontinuity Estimates by District SBAC Threshold Level across Outcomes, Optimal Bandwidth, and Preferred Model Specification

	ITT						TOT					
	Threshold 3		Threshold 2		Difference		Threshold 3		Threshold 2		Difference	
Outcome	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Threshold 3 - Threshold 2												
Std. ELA Score	0.095	0.051	0.02	0.036	0.075	0.062	0.305	0.192	0.118	0.188	0.187	0.269
Std. Math Score	0.027	0.058	0.013	0.067	0.014	0.089	0.139	0.265	0.073	0.346	0.066	0.436
Pr(ELA >= 3)	0.055	0.042	-0.009	0.032	0.064	0.053	0.235	0.18	-0.034	0.163	0.269	0.243
Pr(Math >= 3)	0.003	0.037	0.014	0.028	-0.011	0.046	0.003	0.17	0.075	0.153	-0.072	0.229
Pr(Suspension)	0.013	0.009	-0.008	0.02	0.021	0.022	0.054	0.042	-0.037	0.089	0.091	0.098
Pr(Chronic Absenteeism)	0.014	0.03	0.005	0.038	0.009	0.048	0.082	0.127	0.047	0.165	0.035	0.208

Note. ITT = intent-to-treat. TOT = treatment-on-the-treated. SE = standard error. Std = standardized. ELA = English language arts. Pr = probability. SBAC = Smarter Balanced Assessment Consortium. The preferred model had covariates and district fixed effects (FE) using the calculated optimal bandwidth (specific to each model). p-values: ~: <0.1. *:< 0.05, **: <0.01, ***: <0.001.

As will be addressed in the discussion, these results call into question the conventional framework because they point to a logical inconsistency, i.e. that basic skills thresholds are set too low, but only in the districts that set them high. An alternative hypothesis, as presented in the frameworks section at the beginning of this report, is that reclassification thresholds must be understood in relation to the educational contexts in which they operate. While we lack data to examine differences in instructional environments for EL and reclassified students, we can at compare characteristics of lenient ("SBAC or" and proficiency level 2) versus strenuous ("SBAC only", "SBAC and", and proficiency level 3) districts.

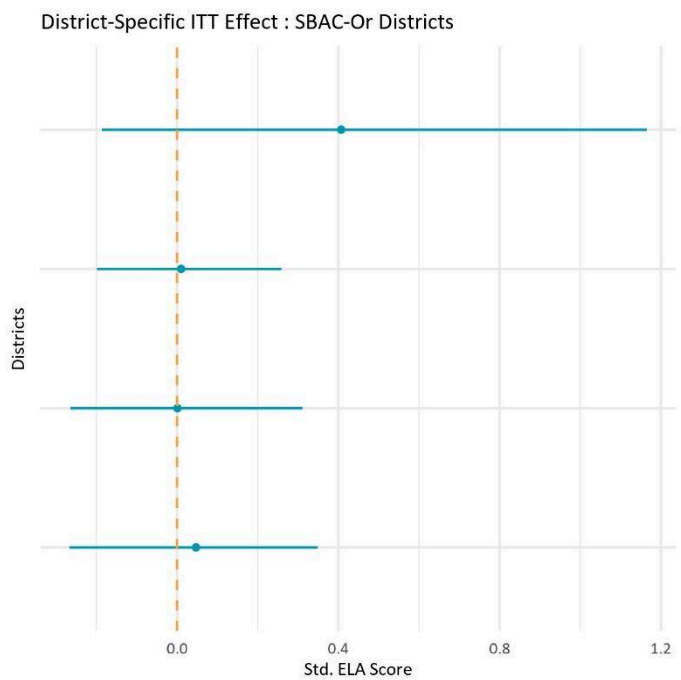
Tables 10 and 12 provide suggestive evidence that the contexts in which students are operating are different in more lenient versus more strenuous districts. Specifically, students in districts with more strenuous basic skills criteria tend to be in far smaller districts, and those serving lower proportions of ever-EL, Latine, and low income students. They also tend to have higher average performance on standardized assessments and lower average chronic absenteeism and suspension rates.

Research Question 5: Among Districts with the Same or Similar Basic Skills Reclassification Criteria, is there Meaningful Heterogeneity by District in the Effects of Reclassification among Students Near the SBAC Threshold?

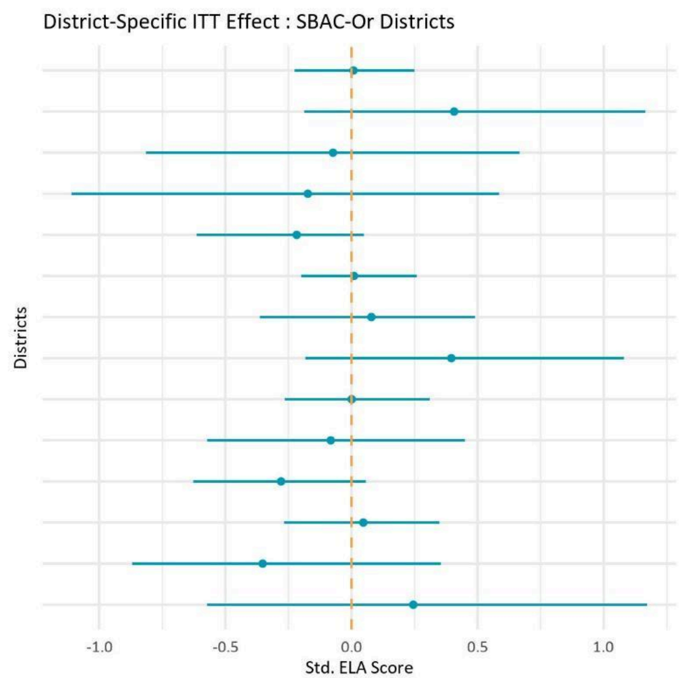
Results from the prior research question cast doubt on conventional frameworks used to understand reclassification criteria threshold levels. Our alternative theory posits that reclassification thresholds must be understood in relation with the opportunities, services, and instructional characteristics that both EL-classified and reclassified students experience. Variation in EL retention effects among districts with the same or similar reclassification criteria can provide at least preliminary evidence to support this hypothesis. This is because variation in effects among same-threshold, same-criteria districts would indicate that effects are not driven solely by threshold or criteria. Our final research question descriptively explores this question, examining heterogeneity in EL retention effects within districts that set their basic skills criteria in the same or similar places. Figures 12-14 present these findings for ELA scale score effects for the "SBAC or", proficiency level 2, and proficiency level 3 subgroups, respectively. Other subgroups and outcomes showed similar variation. As a reminder, power in these district-specific regression discontinuities is limited leading to large standard errors and we limit included districts to those in which reclassification eligibility meaningfully predicted reclassification. We consider this research question exploratory.

Figure 12: District-Specific Intent-to-Treat Estimates of Effect of Being Held in EL Status among English Proficient Students at the Margin of the Basic Skills SBAC Criterion, "SBAC Or" Districts.

A) Restricting to districts where first-stage F-stat > 10



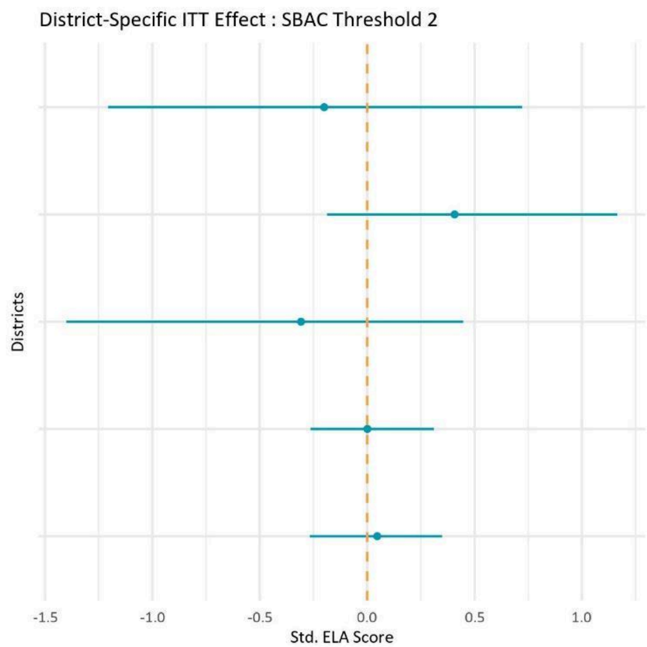
B) Restricting to districts where the jump in the probability of EL-retention was statistically significant ($p < 0.05$)



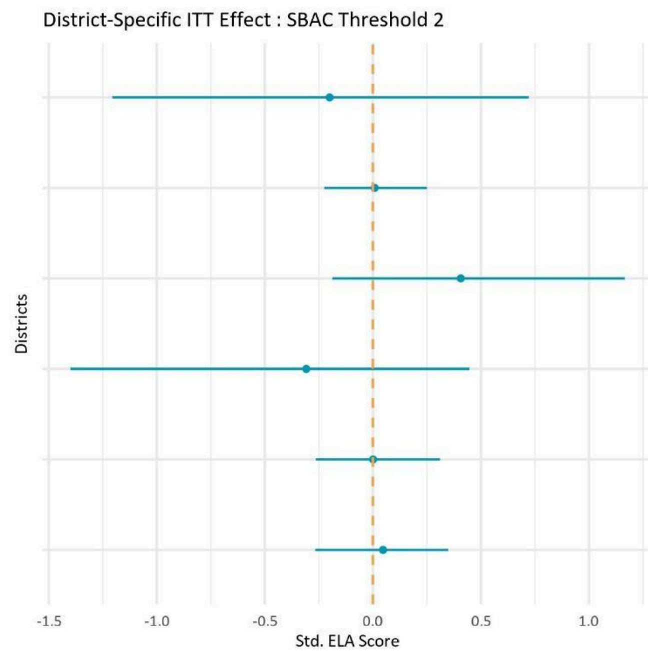
Note. SBAC = Smarter Balanced Assessment Consortium. EL = English learner. ITT = intent-to-treat. Std. = standardized. ELA = English language arts.

Figure 13: District-Specific Intent-to-Treat Estimates of Effect of Being Held in EL Status among English Proficient Students at the Margin of the Basic Skills SBAC Criterion, SBAC Proficiency Level 2 Districts

A) Restricting to districts where first-stage F-stat > 10



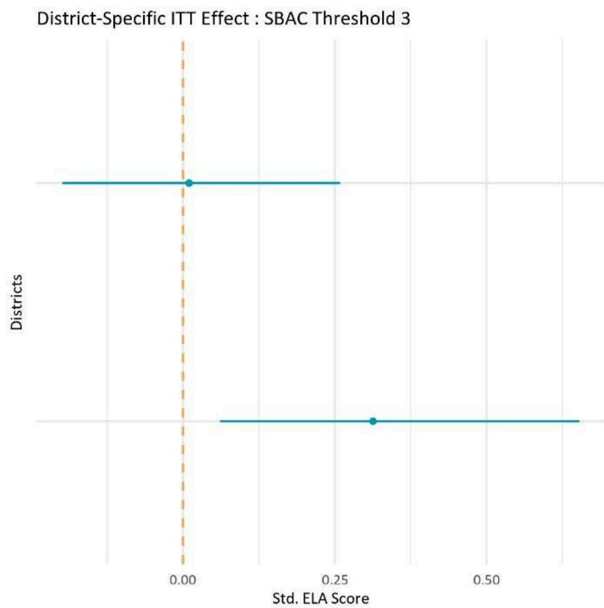
B) Restricting to districts where the jump in the probability of EL-retention was statistically significant ($p < 0.05$)



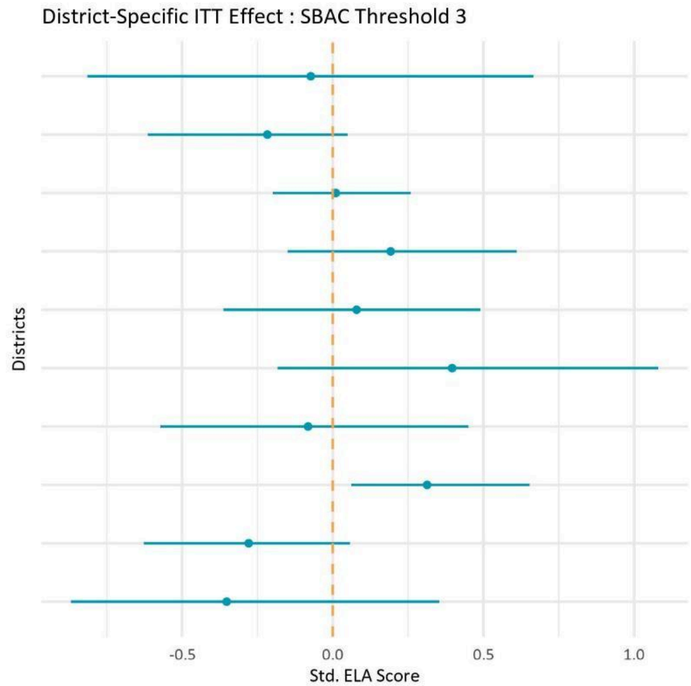
Note. SBAC = Smarter Balanced Assessment Consortium. EL = English learner. ITT = intent-to-treat. Std. = standardized. ELA = English language arts.

Figure 14: District-Specific Intent-to-Treat Estimates of Effect of Being Held in EL Status among English Proficient Students at the Margin of the Basic Skills SBAC Criterion, SBAC Proficiency Level 3 Districts

A) Restricting to districts where first-stage F-stat > 10



B) Restricting to districts where the jump in the probability of EL-retention was statistically significant ($p < 0.05$)



Note. SBAC = Smarter Balanced Assessment Consortium. EL = English learner. ITT = intent-to-treat. Std. = standardized. ELA = English language arts.

Results indicate variation in EL retention effects across districts with the same or similar criteria. Figure 14 for example, shows that among districts that set their SBAC threshold to level 3, EL retention effects on ELA scale scores are both negative and positive, large and small, and statistically significant and non-significant. Similar results are evident in Figures 12 and 13. While not proving our alternative hypothesis, these results indicate that the effects of retaining EL status versus exiting EL status are varied, holding constant thresholds and criteria.

Discussion

While several states rely on more than one measure for English learner reclassification , California is unique in requiring students to satisfy four fully distinct criteria spanning language proficiency, academic performance, teacher evaluation, and parent consultation (Faulkner-Bond, 2026; Morales & Lepper, 2024). As a result, English proficient students may be held in EL status for reasons unrelated to their English language development. This matters because the English learner label, created to protect the educational rights of multilingual students developing English, can also constrain opportunities and experiences in what Alfredo Artiles has called an “equity paradox” (Artiles, 2025). Indeed, a recent study found that while 72% of California’s entering kindergarten cohort reached English proficiency by the end of grade 5, just 50% were reclassified (Novicoff et al., 2025). The rationale for the inclusion of an academic criterion is generally two-fold: first, its inclusion can put pressure on school systems to ensure that EL-classified students are provided with grade appropriate and equitable academic content, and second, its inclusion can ensure that reclassified students are academically prepared to succeed once they exit EL-specific services. But critiques counter that inclusion of an academic criterion is both inappropriate for a language proficiency-related status change and that it potentially penalizes students for failures of school systems to provide strong academic preparation (Linquanti, 2001).

This study set out to learn about how districts are conceptualizing this academic reclassification criterion, how students are faring with it in place, and what impact it has on students who are otherwise reclassification eligible to be held in EL status due to their academic performance. Drawing on a stratified random sample of California districts that closely reflects the composition of the full state, we found that 80% of districts allow students to pass one of multiple exams to clear the basic skills threshold and over 30% set their SBAC proficiency level threshold to 2, below the state mean among English proficient students of 2.53. In this sense, it appears that districts are taking advantage of local control to diminish the impact of the basic skills criterion. This seems to suggest that districts do not want this criterion to be a major impeding factor for students advancing to reclassification, a hypothesis reflected in qualitative research (Mercado-Garcia et al., 2026).

Perhaps because of these flexible or lenient operationalizations of the basic skills criterion, we find that, on average, a considerably larger proportion of students each year meet the basic skills academic criterion compared to the English proficiency criterion. In this sense, districts' efforts to minimize the basic skills criterion as a barrier to reclassification may be working as intended for many students. Yet there is an important subset of districts that are not setting their basic skills criterion low. In our sample, 45% of districts set their SBAC proficiency threshold to 3, well above the state mean, and 6% required students to not only meet the SBAC threshold but to also meet additional test-based measures of basic skills. Further, we find that middle and high school students, and those who attend school in rural districts were more likely than their peers to encounter the basic skills criterion as a barrier. This finding aligns with reclassification timing research out of California that identifies a steep drop off in reclassification rates after elementary school (Umansky & Reardon, 2014). These subgroups of students were more likely to clear ELP thresholds than district SBAC thresholds.

Finally, we also see that there is considerable mismatch between those students that pass the ELP criterion and those that pass the SBAC criterion. Perhaps because these are both high-stake tests that students only can take once per year, we find that about half of students who pass one of the two tests fails to pass the other. This indicates that the two measures may not be well aligned and instead are capturing different sets of knowledge, with a potential downward effect on reclassification eligibility. In sum, despite what appears to be efforts on the part of districts to minimize the basic skills criterion as a barrier to reclassification, it likely remains a barrier for many students.

Despite this, when we look across our sample, we do not find that being held in EL status among English proficient students at the margin of the SBAC criterion has systematic, meaningful, or statistically significant effects on academic or behavioral outcomes. On average, students who are held in EL status do as well as their reclassified peers in the subsequent year. Using a conventional framework for interpreting reclassification effects, a reader would see this finding and deduce that the academic criterion is not only not penalizing students, but that it is placed at precisely the right level at which to facilitate a smooth transition from EL to reclassified status (Betts et al., 2020). But how can this be when districts have such highly varied thresholds and divergent assessments and criteria? How can this be when some districts hold their students to expectations that far exceed state means, while others provide multiple opportunities and set thresholds below the means of demographically similar populations (e.g. FRPL eligible English proficient students)?

In effect, depending on what district a given student attends, the basic skills criterion may be extremely difficult, or quite easy to pass. This variation, found in other studies of California reclassification criteria (Hill et al., 2021), has significant equity implications and led us to explore whether the basic skills criterion is impacting students in different ways across settings. A recent meta-analysis of reclassification effects studies found that the absence of measurable reclassification effects was driven not by true null effects but by wide variation in the directionality and magnitude of reclassification effects across studies (Itoh & Umansky, no date). To examine whether there were systematic differences in reclassification effects by basic skills criterion difficulty we tested whether average effects differed in districts with lenient versus strenuous basic skills criteria.

We found that being held in EL status was not beneficial for students in more lenient districts, a finding that would have indicated that in those districts students were being reclassified too soon, before they were ready to succeed in general education settings. Instead, we found that being held in EL status benefited students in districts with more *strenuous* criteria, a finding the belies conventional framing. How could the criterion be set too low in strenuous districts but just right in lenient districts? These findings do not lend themselves to a simple interpretation of whether California's academic reclassification criterion is helpful or harmful. Instead, they point to the complex relationship between districts' reclassification criteria and other district-specific features that may impact multilingual students' educational outcomes.

Our final research question tested this hypothesis more directly. We tested whether there was meaningful variation in reclassification effects among districts with the same or similar basic skills criteria. Running individual regression discontinuity analyses district by district, we found wide variation in the effects of English proficient students being held in EL status among districts with the same or similar criteria. If the basic skills criterion is operating as intended – ensuring that EL-classified students receive both language supports and full access to academic instruction – then retention in EL status should not produce positive effects in some districts and negative effects in others with the same thresholds. Rather than taking a conventional approach, we instead propose the importance of considering districts' holistic contexts for multilingual students, and specifically the relationship between reclassification criteria and educational opportunity (Cimpian et al., 2017).

Our results point to some relevant policy implications. First, there is no evidence that the academic reclassification criterion uniformly helps or hinders students. Instead, on average, it appears to be neutral indicating that it is likely unnecessary in the California context (Faulkner-Bond, 2026). In other words, we do not find evidence that the academic criterion is, on average, helping by holding students in EL status until better prepared for the general education setting without supports. Instead, our evidence points to many districts strategically setting their academic criterion such that it creates as little of a barrier as possible.

Second, our findings suggest that in districts with harder basic skills criteria there may be a rupture at the time of reclassification. Another study of reclassification in the context of an academic criterion found a similar positive impact of being held in EL status on subsequent ELA scores (Xu et al., 2025). Through mechanism analyses, Xu and colleagues identified that this effect was concentrated in schools that separated ELs from general education classrooms in all subject areas. This highly segregated setting led to abrupt changes, they argued, in services when reclassified. Two other recent studies also identify a short term negative 'shock' of reclassification on student academics, a shock that dissipated and reversed in the years following reclassification (Ma & Winters, 2024; Umansky et al., Under review). Examining the quality and accessibility of instruction and services for both EL and reclassified students in districts where reclassification is associated with academic drops is especially important (de Jong, 2004).

A reclassification shock may emerge under two distinct instructional conditions. On the one hand, as illustrated by Xu and colleagues (2025), EL-classified students may experience diminished teacher expectations and inferior curricular access, leaving them underprepared when they enter more rigorous general education settings (Gándara & Orfield, 2012; Olsen, 2010; Valenzuela, 1999). On the other hand, a reclassification shock could also occur if reclassification results in students losing access to culturally- and linguistically-sustaining instruction where educators understood and valued the assets and skills of multilingual students. In either case, a mismatch between prior instructional context and post-reclassification placement can generate the achievement shock observed in our data.

The fact that we do not see such achievement shocks in all districts (indeed, our sample of districts with more strenuous criteria is a relatively small subset of districts) suggests that an instructional rupture is malleable, and that supports for multilingual students can instead be smoothed and strengthened across educational settings. Examples of smoothing and strengthening include gradual release of language supports for more advanced ELs and ensuring access to important beneficial supports among recently reclassified students. For example, given California's explicit intention to expand bilingual and dual language immersion programs across the state (California Department of Education, 2017, 2018), one approach to maintaining beneficial instructional environments for multilingual students after reclassification is providing students the opportunity to remain in bilingual programs after reclassification. This would not only ensure that these students keep progressing toward full biliteracy and bilingualism, it would also keep them in linguistically- and academically-rich, strength-based instructional settings after being reclassified (Poza, 2021; Steele et al., 2017; Williams et al., 2024). Another important approach, also aligned with state policy (and federal law), is to ensure substantive monitoring of recently reclassified students (de Jong, 2004). Monitoring the opportunities and outcomes of students in their first years in reclassified status can alert educators to students who would benefit from learning supports, and can also ensure identification of students who are not being provided full academic access (California Department of Education, 2019; Umansky et al., 2024).

Finally, although estimates for attendance outcomes were not statistically significant, the point estimates were consistently negative across models, bandwidths, and subgroups, suggesting the possibility that being retained in EL status may be associated with lower attendance. These estimates are imprecise and should therefore be interpreted cautiously, but the pattern warrants further investigation. One possible explanation is that exiting EL status and services could influence students' non-academic experiences at school—for example, by increasing students' sense of belonging or enjoyment in school. Prior qualitative research suggests that EL classification can be experienced as stigmatizing (Dabach, 2014; Valenzuela, 1999), while reclassification may improve aspects of students' well-being and self-efficacy (Lee & Soland, 2022). This dynamic may be particularly relevant for students who have remained classified as ELs for many years, such as long-term ELs, who sometimes describe EL services as insufficiently challenging or demotivating (Kim & García, 2014). Future research is needed to examine whether reclassification influences attendance through these or related mechanisms.

This study has several limitations. First, California has additional non-ELP-related reclassification criteria: teacher approval and parental notification. Our study did not look at these criteria nor how they impact students' trajectories. Research indicates that there is wide variation in the implementation of these criteria as well, suggesting that they, too, may be inequitably creating barriers for English proficient students (Hill et al., 2021). Our impact evaluation also had several limitations. We did not include the whole state, and by necessity we omitted very small districts. As such, our findings do not represent those contexts well. Further, we measured reclassification effects after only one year being reclassified, yet we know that reclassification effects often change over time (Ma & Winters, 2024; Umansky et al., Under review). The first stages of our two stage least squares approach were relatively weak, likely because we didn't have access to student test scores that were not SBAC and ELPAC nor reclassification criteria related to teacher perspective and parental notification. We could only reasonably collect current reclassification criteria in our manual search. As such, we had to assume that these criteria have been in place for a few years in order to conduct our impact evaluation with the most recent data available from CDE. It is likely that at least some of the districts in our sample may have changed their reclassification criteria between 2022-23, the first year of our impact evaluation, and 2025-26, when we collected reclassification data. Relatedly, because we wanted to limit our analysis to only the most recent years of data, we could only examine effects of being held in EL status in the subsequent academic year. Finally, because reclassification criteria are locally determined and we therefore needed to collect them manually, our results apply only to our sample of randomly selected districts.

California stands alone in the complexity and difficulty of its English learner reclassification criteria. As a result, many multilingual students remain EL-classified long after they have reached state definitions of English proficiency. We do not find evidence that this rigorous reclassification policy framework, and in particular, California's requirement that students met academic benchmarks comparable to English proficient peers, benefits students. Instead, we find evidence that many districts attempt to sidestep the basic skills criterion, often with only partial success. Our findings point to the potential benefits of simplifying California's reclassification criteria while focusing increased attention on ensuring full, rigorous, and relevant instruction to California's large and important multilingual population, regardless of language status.

References

Artiles, A. J. (2025). Taming the duality of disability: Critical cultural-historical tools to disrupt equity paradoxes. *Revista Española de Pedagogía, 83*(290), 5-20. https://doi.org/10.22550/2174-0909.4241

Bartlett, C., Cimpian, J. R., & Mavrogordato, M. (2024). Leveling the Playing Field: Default Policy and Its Effects on English Learner Reclassification. EdWorkingPaper No. 24-1085. *Annenberg Institute for School Reform at Brown University*.

Betts, J., Hill, L., Bachofer, K., Hayes, J., Lee, A., & Zau, A. (2020). *Effects of English learner reclassification policies on academic trajectories*.

California Department of Education. (2019). *Monitoring reclassified students*. Retrieved February 26, 2026 from https://www.cde.ca.gov/sp/ml/reclassified122019.asp#:~:text=The%20monitoring%20process%20includes:%20*%20Establishing%20rigorous,language%20proficiency%20(ELP)%20and%20acquiring%20content%20knowledge

California Department of Education. (2025). *Reclassification*. Retrieved February 26, 2026 from https://www.cde.ca.gov/sp/ml/reclassification.asp

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.

Cattaneo, M. D., & Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics, 14*(1), 821-851.

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017, 2017). Evaluating English learner reclassification policy effects across districts. *American Educational Research Journal, 54*(1_suppl), 255S-278S. https://doi.org/https://doi.org/10.3102/0002831216635796

Dabach, D. B. (2014). "I am not a shelter!" Stigma and social boundaries in teachers' accounts of students' experience in separate "sheltered" English searner classrooms. *Journal of Education for Students Placed at Risk, 19*(2), 98-124.

de Jong, E. J. (2004). After exit: Academic achievement patterns of former English language learners. *education policy analysis archives, 12*.

Estrada, P., & Wang, H. (2013). Reclassifying and not reclassifying English learners to fluent English proficient, year 1 findings: Factors impeding and facilitating reclassification and access to the core. American Educational Research Association Annual Conference, San Francisco, California

Every Student Succeeds Act, ESSA (2015).

Faulkner-Bond, M. (2026). *Finding the sweet spot for English learner reclassification: A study of California's criteria*.

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review, 85*(2), 149-171.

Gándara, P., & Orfield, G. (2012). Segregating Arizona's English learners: A return to the "Mexican room"? *Teachers College Record, 114*(9), 1-27.

Garrett, R., Larsen, E., & Arellanes, M. (2022). *Effects of reclassifying English learner students on student achievement in New Mexico* [REL](2022–138).

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*(4), 328-331.

Hill, L., Lee, A., & Hayes, J. (2021). Surveying the landscape of California's English learner reclassification policy. *Public Policy Institute of California*, 1-21.

Imbens, G. W., & Lemieux, T. (2008, 2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics, 142*(2), 615-635.

Itoh, T., & Umansky, I. (no date). How does exiting English learner status impact multilingual students' achievement and attainment? *HEDCO Institute for Evidence-Based Educational Practice*. https://hedcoinstitute.uoregon.edu/blog/11/english-learner-reclassification

Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher, 43*(8), 381-389.

Johnson, A. (2019). The effects of English learner classification on high school graduation and college attendance. *AERA Open, 5*(2).

Johnson, A. (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/https://doi.org/10.3102/0162373719877197

Kim, W. G., & García, S. B. (2014). Long-term English language learners' perceptions of their language and academic learning experiences. *Remedial and Special Education, 35*(5), 300-312.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241-253.

Lee, M. G., & Soland, J. G. (2022). Does reclassification change how English learners feel about school and themselves? Evidence from a regression discontinuity design. *Educational Evaluation and Policy Analysis*, 01623737221097419.

Ma, M., & Winters, M. A. (2024). *Reclassifying English learners* (EdWorkingPaper, Issue 25-1234). https://edworkingpapers.com/sites/default/files/ai25-1234.pdf

Mavrogordato, M., & White, R. S. (2017). Reclassification variation: How policy implementation guides the process of exiting students from English learner status. *Educational Evaluation and Policy Analysis, 39*(2), 281-310.

Mercado-Garcia, D., Sipes, L., & Solano-Flores, G. (2026). *Creating coherence: Lessons on English learner reclassification* Policy Analysis for California Education webinar, https://edpolicyinca.org/events/rethinking-english-learner-reclassification-californi

Morales, C., & Lepper, A. (2024). The changing landscape of states' English learner reclassification policies. *Center for Education Efficacy, Excellence, and Equity (E4), Northwestern University*.

Novicoff, S., Reardon, S. F., & Johnson, R. (2025). A longitudinal portrait of California's kindergarten English learners & their learning outcomes. *AERA Open, 11*.

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

Onda, M., & Seyler, E. (2020). English learners reclassification and academic achievement: Evidence from Minnesota. *Economics of Education Review, 79*, 102043.

Poza, L. E. (2021). Adding flesh to the bones: Dignity frames for English learner education. *Harvard Educational Review, 91*(4), 482-510.

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness, 5*(1), 83-104.

Robinson, J. P. (2011, 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*(3), 267-292. https://doi.org/https://doi.org/10.3102/0162373711407912

Robinson-Cimpian, J. P., & Thompson, K. D. (2016, 2016). The effects of changing test-based policies for reclassifying English learners. *Journal of Policy Analysis and Management, 35*(2), 279-305. https://doi.org/https://doi.org/10.1002/pam.21882

Smarter Balanced Assessment Consortium. (2026). *Percentiles*. Retrieved February 26, 2026 from https://validity.smarterbalanced.org/reports-and-specifications/percentiles/

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica, 65*(3), 557-586.

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement: Evidence from lottery data. *American Educational Research Journal, 54*(1_suppl), 282S-306S.

Thompson, K. D., Umansky, I. M., & Rew, W. J. (2023). Improving understanding of English learner education through an expanded analytic framework. *Educational Policy, 37*(5), 1315-1348.

U.S. Department of Justice & U.S. Department of Education. (2015). *Dear colleague*.

Umansky, I., Itoh, T., Faulkner-Bond, M., Porter, L., & Tanner-Smith, E. (Under review). English learner reclassification as an educational input: Meta-analytic evidence of the impact of existing English learner status.

Umansky, I. M., & Reardon, S. F. (2014, 2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal, 51*(5), 879-912.

Umansky, I. M., Shin, N., Thompson, K. D., Avelar, J. D., & Bovee, J. (2024). Beyond English instruction: A multistate analysis of Lau's unfulfilled promise of access to core content. *Bilingual Research Journal, 47*(4), 504-531.

Valdés, G. (2005). Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal, 89*(3), 410-426.

Valenzuela, A. (1999). *Subtractive schooling: US-Mexican youth and the politics of caring*. State University of New York Press.

Williams, C., Umansky, I., Porter, L., Vazquez Cano, M., & Zabala, J. (2024). *Meeting its potential: A call and guide for universal access to bilingual education in California*.

Williams, C., & Zabala, J. (2023). *Moving from vision to reality: Establishing California as a national bilingual education and dual-language immersion leaders*.

https://tcf.org/content/report/moving-from-vision-to-reality-establishing-california-as-a-national-bilingual-education-and-dual-language-immersion-leader/

Xu, S., Flanagan, C., & Quintero, D. (2025). *Bulwark or barrier? The effect of academic criteria-based reclassification on the high school outcomes of multilingual learners in Texas. EdWorkingPaper No. 25-1368*.