Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1368

Bulwark or Barrier? The Effect of Academic Criteria-Based Reclassification on the High School Outcomes of Multilingual Learners in Texas

Shirley Xu
Vanderbilt University

Coral Flanagan
Vanderbilt University

Diana Quintero
Brown University

English learner (EL) classification can provide multilingual students (MLs) with key supports while simultaneously limiting access to important educational opportunities. To determine when students are ready to exit EL status, some states require students to meet academic criteria in addition to demonstrated English proficiency. However, few studies empirically examine these criteria, which may become increasingly restrictive in upper grades. We use administrative data from Texas and a regression discontinuity design to estimate the effect of using academic criteria to reclassify students at the end of eighth grade. Reclassified students earn more advanced credits, but graduate high school at lower rates, indicating a need for greater course access for EL-classified students and increased support for former ML-ELs.

VERSION: December 2025

Suggested citation: Xu, Shirley, Coral Flanagan, and Diana Quintero. (2025). Bulwark or Barrier? The Effect of Academic Criteria-Based Reclassification on the High School Outcomes of Multilingual Learners in Texas. (EdWorkingPaper: 25-1368). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/9hwj-cs28

Bulwark or Barrier? The Effect of Academic Criteria-Based Reclassification on the High School Outcomes of Multilingual Learners in Texas

Shirley Xu[1]
Vanderbilt University
shirley.h.xu@vanderbilt.edu

Coral Flanagan[2]
Vanderbilt University
coral.flanagan@vanderbilt.edu

Diana Quintero[3]
Brown University
diana_quintero_castellanos@brown.edu

Acknowledgements

The authors thank the Texas Education Agency (TEA) and the University of Houston Education Research Center (UH ERC) for access to the data used in this study. We also thank the UH ERC staff who provided crucial technical support for this project. We are grateful for the thoughtful feedback and comments from Shaun Dougherty and discussants and session participants at the Fall 2023 meeting of the Association for Public Policy Analysis and Management, the Spring 2024 meeting of the Association for Education Finance and Policy, and the Peabody Education Policy Seminar Series at Vanderbilt University.

[1] Corresponding author
[2] Equal second authorship
[3] Equal second authorship

***

English learner (EL) classification can provide multilingual students (MLs) with key supports while simultaneously limiting access to important educational opportunities. To determine when students are ready to exit EL status, some states require students to meet *academic* criteria in addition to demonstrated English proficiency. However, few studies empirically examine these criteria, which may become increasingly restrictive in upper grades. We use administrative data from Texas and a regression discontinuity design to estimate the effect of using academic criteria to reclassify students at the end of eighth grade. Reclassified students earn more advanced credits, but graduate high school at lower rates, indicating a need for greater course access for EL-classified students and increased support for former ML-ELs.

***

Multilingual students classified as "English learners" (EL)—hereafter "ML-ELs"[1]—are one of the fastest growing groups in the United States, composing over 10% of all public school enrollment (Hussar et al., 2020). To ensure equitable educational experiences and outcomes, federal mandate requires states to develop policies for formally identifying students as ELs, supporting them toward attaining English proficiency, and ultimately reclassifying them as "English proficient" (Every Student Succeeds Act [ESSA], 2015). EL classification can provide students with critical supports, including English language development (ELD) instruction (Baker et al., 2014), home-language content instruction (Porter et al., 2023; Steele et al., 2017), and specialized teacher expertise (Master et al., 2016). However, retaining students in EL programming after support services are no longer needed may become actively detrimental, given that EL status is also linked to linguistic isolation (Gifford & Valdés, 2006), assignment to less experienced or effective teachers (Gándara, 2021; Torre Gibney & Henry, 2020), and tracking into lower-level classes (Dabach, 2014; Estrada, 2014; Kanno & Kangas, 2014). Accurately determining the right time for reclassification is therefore a critical challenge for policymakers.

Each state sets their own reclassification criteria, typically combining English language proficiency (ELP) assessments with additional indicators such as teacher evaluations (Rafa et al., 2020). While a growing body of research exploits test-based criteria to measure the causal impact of reclassification on subsequent outcomes, far less is known about how the inclusion of particular criteria in the reclassification process shapes those outcomes (Morales & Lepper, 2024). As of 2023, 18 states include non-ELP assessments in the reclassification process. Among them, Texas, Florida, and New York—three of the states serving the largest numbers of ML-ELs—require students to meet minimum scores on statewide English language arts (ELA) content exams in order to reclassify. Other states like Nevada and South Dakota include ELA assessments as alternative criteria for reclassification for students who have lower ELP scores. The use of ELA content exams in reclassification decisions remains contentious: some scholars contend that requiring students to demonstrate academic content mastery may unnecessarily delay reclassification for otherwise proficient students (Abedi, 2008; Umansky & Porter, 2020). However, empirical evidence on the impacts of including academic content exams in the reclassification process is sparse.

In this study, we provide some of the first state-level evidence on the effects of using academic reclassification criteria. Drawing on administrative data from Texas, we use a frontier regression discontinuity (RD) design in which we first condition our sample on meeting all ELP exam thresholds and then estimate the effect of just barely passing the reclassification threshold on the state's eighth-grade ELA content assessment. Our study is also one of the first to study reclassification in Texas—the state that serves the second largest population of EL-classified students in the U.S.—bringing rigorous evidence to bear on current debates about the appropriateness of using academic reclassification criteria and the role of the state in setting policy that ensures educational thriving for all students (Texas Education Agency [TEA], 2024; Umansky & Porter, 2020). Finally, as students progress through school, the likelihood of reclassification decreases while course differentiation and tracking increase (Gamoran, 2009; Johnson, 2020; Long & Iatarola, 2012). EL status in high school may thus meaningfully affect students' instructional environment and subsequent outcomes. We extend the relatively sparse literature on reclassification in later grades by focusing specifically on reclassification in eighth grade, during the critical transition into high school (Johnson, 2019).

For eighth-grade students with advanced English proficiency (those who met all three English language requirements for reclassification), we find that reclassifying based on ELA criteria has a significantly positive impact on the likelihood of earning advanced course credit in high school. However, we also observe large negative impacts on the likelihood of graduating high school, with suggestive evidence that reclassified students were more likely to fail core academic courses in high school and less likely to pass the Algebra I end-of-course exam, a requirement for high school graduation. Our results are robust to different model specifications and bandwidths. Taken together, our findings indicate that using ELA content exams to reclassify ML-EL students may have both positive and negative impacts on the marginal student's high school outcomes. Our mixed results do not offer conclusive evidence on how states should incorporate academic criteria into the reclassification process. However, our findings do suggest that ML-EL students who demonstrate potential should be supported in accessing advanced courses, regardless of language status. Simultaneously, policymakers and practitioners should carefully monitor reclassified students and increase support for those who struggle in the period following program exit.

Conceptual Framework

The Potential Effects of EL (Re)Classification

Under federal policy, public schools are required to identify students whose English proficiency level would preclude full participation in English-only instructional settings as "English learners" and provide them with services to support both English language and content learning (ESSA, 2015). Classification as an EL can alter various aspects of students' instructional environments, including curriculum, course placement, teachers, and peer composition (Umansky & Porter, 2020). Prior literature suggests that staying EL-classified can both benefit and harm students, depending on their readiness to succeed in mainstream classes and how classification changes students' instructional settings (e.g., school services, resources) in each specific context.

EL classification can benefit students by providing access to additional services and supports including direct instruction in English language skills (Hopkins et al., 2022)), bilingual instruction (Porter et al., 2023), and English-as-a-second-language (ESL)- or bilingual-certified teachers (Loeb et al., 2014; Master et al., 2016). When implemented well, these resources can bolster both academic and English language proficiency (Master et al., 2016; Porter et al., 2023). Students typically lose access to these services when they exit EL classification, so reclassifying too early can negatively affect outcomes by removing necessary supports. For example, in their study of the Los Angeles Unified School District (LAUSD), Robinson-Cimpian and Thompson (2016) found that EL-classified students who just barely met criteria to reclassify in Grades 9 or 10 had lower ELA performance and were less likely to graduate from high school than similarly achieving peers who remained EL-classified. These effects disappeared in subsequent years after policymakers adopted more stringent reclassification criteria, which suggests that prior negative effects were driven by students losing access to beneficial supports before they were ready.

Conversely, EL classification may negatively impact other aspects of students' educational experiences. A growing body of literature finds that EL status is associated with decreased access to educational resources such as highly experienced teachers, rigorous courses, and exposure to core academic content (Callahan et al., 2010; Dabach, 2014; Gándara et al., 2003; Torre Gibney & Henry, 2020; Umansky, 2016a). Policies that prioritize English acquisition at the expense of exposure to academic content knowledge may contribute to these disparities (Umansky & Avelar, 2023). For example, many schools meet federal requirements to support English acquisition by enrolling students in one or multiple English language development (ELD) courses during the school day. In some contexts, ELD can crowd out space in students' schedules for other classes, including core academic courses (Lillie et al., 2012; Umansky, 2016a). Another common strategy for supporting both academic and language proficiency development is to place ML-ELs in EL-specific content courses that often offer less rigorous instruction and lower academic content coverage (Dabach, 2014). Further, research suggests that formally labelling students as English deficient can be academically and socially stigmatizing (Umansky, 2016b). Research shows that teachers tend to view EL-classified students as less academically capable than non-EL peers with comparable test scores (Umansky & Dumont, 2021) and that students who remain classified as EL report lower academic self-confidence than those who narrowly reclassify (Lee & Soland, 2022). Positive effects of reclassification on outcomes like achievement may, therefore, signal that EL status imposes harmful constraints on learning that are removed when students exit, or that students are being retained for too long in EL classification.

Finally, exiting EL status may *simultaneously* harm and benefit students. For example, reclassified students may lose supports like testing accommodations and helpful instructional scaffolding; at the same time, they may gain access to more college preparatory content. The observed impacts of reclassification thus represent the combined effect of losing supports *and* removing constraints, with the direction of results depending on the magnitude of harm versus benefit. Given that the particular circumstances surrounding ML-ELs mediate reclassification effects on subsequent student outcomes, it is important to provide evidence about the impacts of reclassification from multiple contexts.

The Use of Academic Content Criteria

All states use some measure of English language proficiency to determine reclassification but some apply additional criteria; for instance, several include performance on ELA content exams in reclassification decisions (Morales & Lepper, 2024; Rafa et al., 2020). Although the specific assessments used in each state vary, ELA content assessments are designed to measure different constructs than ELP reading and writing assessments; this difference is particularly pronounced in later grades. While ELP assessments often focus on decoding, vocabulary, understanding grammatical structures, and analyzing basic information and ideas in texts, ELA assessments typically assess critical reading skills like understanding abstract language, inferring authorial intent, and comprehending literary or scientific texts (Forte et al., 2012). Proponents argue that including ELA achievement measures ensures that ML-EL students are academically prepared to succeed in a mainstream classroom environment, acting as a bulwark against early exit (Linquanti, 2001; Ragan & Lesaux, 2006). Critics counter that state content exams are not designed to assess language learners and thus produce biased test scores that fail to represent ML-ELs' true content knowledge (Umansky & Porter, 2020; Wolf & and Leon, 2009). Further, English-only students are typically not required to meet state thresholds on ELA content exams in order to be placed in mainstream classrooms, which creates policy incoherence (Abedi, 2006; Bailey & Carroll, 2015; Gándara & Baca, 2008; Solano-Flores, 2008; Umansky & Avelar, 2023).

Concern over the growing population of long-term ML-ELs is one factor fueling criticism of using academic criteria for reclassification. Research suggests that it typically takes up to seven years to develop proficiency in a new language (Thompson, 2017), yet many students remain EL-classified for much longer, with one meta-analysis finding that nearly a quarter of students who started kindergarten as ELs never reclassified (Garibay & Abdelkader, 2024;

Olsen, 2010; Thompson, 2017). Because reading and writing proficiency criteria are usually the hardest for students to meet, particularly at the secondary level, critics suggest that including ELA content exams in the reclassification process increases the likelihood that students will be retained in EL status for multiple years based on a single criterion (Abedi, 2006; Robinson, 2011; Thompson, 2017). Remaining in EL services that cater to lower English proficiency students can initiate a vicious cycle wherein students feel stigmatized and lose confidence, disengage from school, and consequently demonstrate decreased academic performance—that is then used to hold them in EL status (N. Flores et al., 2015; Menken et al., 2012; Umansky & Avelar, 2023). However, empirical evidence on the utility of using content exams for reclassification is limited. Reclassification effects are often assessed through regression discontinuity studies that apply a binding-score design and thus reduce multiple reclassification criteria to a single composite score (Johnson, 2019; Reyes & Hwang, 2021; Robinson, 2011). Very few studies systematically evaluate which tests contribute to reclassification, and whether there are heterogeneous effects by threshold.

Reclassification Timing and the Middle-to-High-School Transition

The relative benefits and harms of reclassification may also depend in part on its timing. Researchers frequently find that reclassification effects vary across grade levels or bands (Betts et al., 2020; Pope, 2016; Robinson, 2011) and may be strongest during the transition between school-levels (Johnson, 2019). The evidence base on reclassification at the secondary level is generally more limited than its elementary equivalent. However, a growing body of research examines reclassification effects in middle or high school. Each of the three potential directions of reclassification effects discussed above bears out in these studies, which find a range of positive, negative, and null impacts on measures of math and ELA achievement in several different contexts (Betts et al., 2020; Carlson & Knowles, 2016; Cimpian et al., 2017; Johnson, 2020; Pope, 2016; Reyes & Hwang, 2021; Robinson, 2011; Robinson-Cimpian & Thompson, 2016).

Research on the impacts of later grade reclassification on high school graduation is similarly mixed. In Wisconsin, Carlson and Knowles (2016) find suggestive evidence that reclassifying in 10th grade increased high school graduation rates and stronger evidence that it increased students' likelihoods of postsecondary enrollment. In another state, however, reclassification in middle and high school had no effect on graduation (Cimpian et al., 2017). Two studies found that high school reclassification in the Los Angeles Unified School District had a negative effect on graduation that disappeared when the criteria became more stringent (Betts et al., 2020; Robinson-Cimpian & Thompson, 2016), while Johnson (2019) found null effects of eighth-grade reclassification on graduation rates in another California district. Altogether, the extant literature on secondary reclassification might best be described as conclusively varied.

It is not surprising that effects of reclassification differ across studies, given that both reclassification criteria and English learner services vary widely across state and even district lines. For example, in their study of eighth-grade reclassification in a California district, Reyes and Hwang (2021) attribute null findings to the lack of variation in instructional settings for ML-ELs and mainstream students in their study context. In contrast, Carlson and Knowles (2016) suggest that 10th-grade reclassification may be particularly influential because it influences ELA curriculum exposure in 11th grade, the year that all students in Wisconsin take the ACT. In the absence of automatic reclassification processes (Bartlett et al., 2024), school districts may also vary in how strictly or loosely they follow state guidelines. Cimpian, Thompson, and Makowski

(2017) directly examined variation in the degree of reclassification policy adherence among districts in two different states and found significant heterogeneity in both the likelihood of reclassification and its subsequent effects on student achievement and graduation. These studies emphasize the importance of context and highlight the need for more research on the effects of reclassification in different settings.

As described above, reclassification may have heterogeneous effects across not only contexts, but also *outcomes*, varying in both direction and magnitude. However, existing evidence on some key outcomes is sparse. For example, disparity in access to academic content is one of the most frequently cited potential negative effects of EL classification. While a large body of descriptive research suggests that ML-ELs are less likely than peers to take advanced or upper-level courses in middle and high school (Callahan et al., 2010; Cashiola et al., 2022; Johnson, 2019; Zuniga et al., 2005), only three regression discontinuity studies address this topic, all suggesting that EL classification has no impact on advanced course engagement (Reyes & Hwang, 2021; Robinson, 2011; Umansky, 2016b). There remains a need for more research on course access as well as other outcomes that shape students' in-school experiences.

Taken together, the existing base of literature on reclassification suggests that exiting EL status can be consequential for students' future success in certain contexts, but research on the specific impacts of eighth-grade reclassification and on the efficacy of using academic content criteria in the process is currently limited. In addition, the majority of past studies are single-district cases situated in California. Our study extends the small body of work that examines the impacts of reclassification during the critical juncture between middle and high school by studying academic-based reclassification in Texas, providing new causal evidence on how EL classification impacts course taking, achievement, and attainment.

State Policy Context

We situate our study in Texas, the state that currently has the highest percentage of EL-classified students, at nearly a fifth of all public-school enrollment (National Center for Education Statistics [NCES], 2024). Ninety-nine percent of school districts in Texas serve EL-identified students; the concentration of ML-ELs is highest in districts along the Texas-Mexico border in the Rio Grande Valley and El Paso regions (31–40%) and in the regions surrounding the major urban centers of Houston, Austin, and Dallas (21–30%) (TEA, 2024). The most common language spoken by ML-EL students in the state is Spanish by a large margin—over 85%, followed by Vietnamese and Arabic at less than 2% each.

Texas Reclassification Requirements

School- or district-level Language Proficiency Assessment Committees (LPACs) in Texas make reclassification determinations each spring based on exit criteria provided by the Texas Education Agency. During our study period from 2013-2017, TEA stipulated that reclassification decisions for eighth-grade ML-ELs should be based on the following criteria: a) English oral (i.e., speaking and listening) and writing proficiency, as measured by scoring "advanced high" on corresponding domains on the Texas English Language Proficiency Assessment System (TELPAS) or fluent on a different assessment from a state-approved list;[2] b) English reading proficiency, as measured by meeting the passing standard on the State of Texas Assessments of Academic Readiness (STAAR), the state ELA content assessment; and c) a subjective teacher evaluation. Notably, although federal guidance recommends using a "valid and reliable English language proficiency (ELP) assessment" for all four domains, Texas did not use an ELP reading test for reclassification during our study period and used only the state ELA content exam to assess reading proficiency (U.S. Department of Education, (2015), p. 3).

Replacing ELP reading with the state context exam could be consequential for students along two dimensions: 1) differences in content assessed; 2) differences in relative difficulty of obtaining the passing threshold. In contrast to ELP reading tests which typically assess basic comprehension, ELA tests assess critical reading skills. Appendix B, Figure B1 presents sample items from Texas' ELP reading exam (not used for reclassification) and ELA content exam. TELPAS (ELP) asks students to complete a sentence with the correct vocabulary word; STAAR (ELA) asks students to identify evidence in a three-paragraph article. It bears noting, however, that the passing standard for STAAR was also relatively lower than the standard for TELPAS. During our study period, 44% of eighth-grade ML-ELs passed STAAR Reading on the first try, while only 24% attained the required standard of "advanced high" on TELPAS reading. Therefore, Texas' choice to replace ELP reading with the ELA exam both emphasized critical reading skills and held students to a relatively lower standard for demonstrating reading proficiency to reclassify.

English Learner Services in Texas

At the secondary level, most ML-ELs in Texas are enrolled in ESL courses, in which educators teach in English while providing instructional supports to foster both English language acquisition and academic proficiency (Appendix A, Table A1). Schools in Texas can choose to offer ESL services either in all core content courses or only in ELA (TEA, 2025, §89.1201).[3] ESL services can be offered in either separate classes or integrated classes with both ELs and non-ELs (Dabach, 2014; Gifford & Valdés, 2006).

Unique Characteristics of the Texas Context

A few characteristics of the Texas context distinguish it from other settings of prior reclassification studies. First, Texas state law requires all elementary schools that enroll 20 or more ML-ELs to offer bilingual instruction in stark contrast to states like California where bilingual education was legally restricted until 2016. This early access to bilingual programs may have implications for students' readiness to succeed in mainstream courses when reclassifying in eighth grade. Second, the central role of the LPAC in Texas reclassification is fairly unique. Other states like California and Wisconsin do not require the establishment of local committees; in fact, reclassification in Wisconsin is automated at the state level (Carlson & Knowles, 2016). Situating this analysis in the relatively decentralized Texas context may also provide some transferable insights to states like Florida, which also serves large numbers of ML-ELs and involves committee-based decisions making (Florida Administrative Code, Rule 6A-6.0903).

Lastly, Texas requires all students, including ML-ELs, to pass five end-of-course (EOC) exams in ELA, math, science, and social studies to graduate (Texas Administrative Code, §101.3022), creating a set of high stakes consequences that may come to define large swathes of the high school educational experience. In comparison to states that lack such extensive testing requirements (e.g., California and Wisconsin), graduate test requirements may potentially widen the gap between EL and mainstream instructional settings. Under these circumstances, holding students past the point at which they stop benefitting from EL programming or exiting them prematurely could be particularly detrimental to their learning and ultimate K-12 attainment outcomes.

Research Design and Methods

Sample

This study draws upon administrative data from Texas public schools. We focus on five cohorts of students from the 2012-2013 to 2016-2017 academic years who were classified as English Learners in eighth grade and expected to graduate high school between 2017 and 2021.

We restrict our sample to students who: 1) were EL-classified in eighth grade; 2) had valid test scores on both TELPAS and STAAR reading; 3) scored "advanced high" in three TELPAS language domains: speaking, listening, and writing (32% of all eighth-grade ML-ELs); 4) took the second administration of STAAR; and 5) re-enrolled in a public school in the state of Texas in ninth grade. The second and third restrictions are necessary for the frontier RD design at the core of our analysis; the fifth allows us to identify students' EL-status in the following year.

The fourth restriction creates a cleaner comparison between reclassifiers and those who remain EL-classified (Hwang & Koedel, 2023). All students in Texas take STAAR at least once in the spring of eighth grade (in March or April). Almost all students who fail on their first attempt retake the exam in May or June, and ML-ELs who pass on the second try become eligible to reclassify.[4] A large number of ML-ELs who score below the passing threshold on the first attempt exceed it on the second and go on to reclassify. Using the second STAAR administration allows us to compare students who ultimately qualify for reclassification to similar peers who never met STAAR criteria, creating a clear contrast to better isolate the treatment effect. Table A2 in Appendix A shows that relative to peers who take STAAR on the first attempt, students who take the second attempt are lower performing on metrics like seventh grade achievement (See Columns 3 and 4). Our findings thus measure the impacts of exiting EL-status for a subgroup of students with relatively high English language proficiency and more marginal academic proficiency, a common profile among secondary ML-ELs. Our final analytic sample includes 21,044 observations (Table A2, Column 6).

[TABLE 1]

Table 1 provides descriptive characteristics for our sample (Column 2) vis-à-vis all eighth-grade ML-ELs (Column 1). Our analytic sample is similar to the entire population of eighth-grade ML-ELs across observable characteristics, with three key differences. First, 95% of students in the analytic sample have been in US schools for at least 5 years compared to 72% of all eighth-grade ML-ELs. Second, the representation of students receiving special education is slightly smaller in our sample (12% of ML-ELs vs. 7% of students in our sample). Third, as we discussed above, we choose to use students who take the second administration of STAAR, typically because they failed the first administration, as our analytic sample. Thus, our sample has notably lower seventh-grade achievement than the full eighth-grade ML-EL population. It is mostly composed of long-term ML-ELs who have relatively high English language proficiency and lower ELA proficiency, a group that may be particularly negatively affected by including ELA content exams in the reclassification process (Morales & Lepper, 2024; Olsen, 2010).

Measures

We consider the relationship between ELA content-based reclassification in eighth grade and three sets of high school outcomes: course taking, academic achievement, and graduation. For the remainder of the article, we simply use “reclassification” for brevity; however, we are always referring to reclassification under the timing and criteria conditions specified above.

Treatment Variable

Eighth-grade students in Texas were assigned a raw score on the STAAR Reading test ranging from 0 to 52 (scores ranged from 0 to 44 in academic year 2016-17 only) and a corresponding scale score ranging roughly from 900 to 2200. To achieve passing standard, students’ scaled scores must fall above the threshold for “Satisfactory” performance (“Approaches standard” in 2016-17). Because the scale scores increase discontinuously and vary in their discrete values across years, we utilize the raw scores for the running variable, using score conversion tables provided by TEA to identify the passing cutoff in each year.[5] The treatment variable for our reduced-form, intent-to-treat model is a binary indicator equal to 1 if a student scored at or above the passing threshold and achieved *eligibility* for reclassification, and 0 if they scored below the threshold. However, because final reclassification decisions in Texas are determined by the human judgment of the LPAC, not all students who score above this cutoff are necessarily reclassified, or "treated." We discuss our empirical strategy for addressing this fuzziness in the reclassification process in greater detail below.

It is also important to note that EL classification is a dynamic status. Our analyses compare students who barely meet the criteria to reclassify in eighth grade to those who just miss the threshold and thus remain EL-classified. However, Table A3 (Appendix A) demonstrates that a portion of students in the comparison group eventually reclassify in subsequent years. Our research should thus be thought of as measuring the impact of *entering* high school classified as English proficient rather than as an English learner on subsequent outcomes in high school.

Outcome Variables

High School Course Taking. We consider the impact of reclassification on the likelihood of earning credit in two types of high school courses. First, we measure the likelihood of earning credit by earning a passing grade in an "advanced" core content course, which we operationalize as either Advanced Placement (AP), International Baccalaureate (IB), or Dual Enrollment (DE). We focus on these three course types, specifically, to measure access to upper-level courses that are designed to prepare students for college and to opportunities to earn college credit, whether through exams or partnerships with postsecondary institutions. We consider whether students earn any AP, IB, or DE credit, broadly, and in each core content area, and whether they earn multiple credits, which can indicate increased college readiness and a financial advantage if they pursue postsecondary education.

Second, we consider the likelihood of enrolling in and completing (i.e., earning credit in) remedial or non-college preparatory STEM classes. These include remedial math (defined as either a “Math Models” or “Strategies for High School Math” course) and an integrated physics and chemistry (IPC) course which replaces the more rigorous separate courses that are recommended for admission to the University of Texas at Austin—the state’s flagship university. If we observe an increased likelihood of taking advanced courses or decreased likelihood of taking remedial/non-college prep courses, this might suggest that keeping students EL-classified based on their STAAR Reading achievement constrains course access. All course taking outcomes are measured using binary indicator variables.

High School Achievement. We consider three measures of high school achievement. First, we consider performance on STAAR EOC exams, designed to assess students’ knowledge and skills in specific content areas. During our study period, Texas required students to pass five EOC exams to graduate high school: Algebra I (typically taken in Grade 9), English I (Grade 9), Biology (Grade 9), English II (Grade 10), and US History (Grade 11). Students who fail on the first attempt can retake the tests or prove proficiency using an alternative test like the SAT. It is important to note that Texas allows students to retake an EOC exam without retaking the corresponding course if they achieved a passing grade on the first attempt. We first measure the impact of reclassification on students’ standardized scores on each exam, which provides some indication of their content proficiency, regardless of pass-fail status. Using a binary outcome variable, we also examine the likelihood of having to retake each EOC, which can shape subsequent course access and hinder graduation.

Given the limitations of standardized tests when it comes to measuring student academic performance, we consider an alternative measure: the likelihood of ever failing a course in one of the four core academic subjects during high school. This serves as a rough proxy for student grades, which we do not have access to. It may also be a function of other constructs our data preclude us from examining, such as students' affective engagement with school. If we observe negative reclassification effects on achievement, this could suggest that students who are relatively proficient in English are nevertheless losing access to necessary supports too early—or that students who remain EL are being placed in less rigorous courses. Conversely, evidence that reclassification improves performance might suggest that exiting EL status increases access to rigorous instruction and improves exposure to academic content (Dabach, 2014).

High School Graduation. Finally, we consider four graduation outcomes: 1) the likelihood of ever graduating high school by Spring of 2022, our last year of data; 2) the likelihood of graduating high school on time, defined as in spring of the academic year that is four years after the student first enrolled in eighth grade; 3) the likelihood of graduating college ready, which we define as taking four English credits, three math credits (including Algebra I and II), three science credits (biology, chemistry, and physics), three social studies credits, and two world language credits, based on the recommended course sequence for admission to UT-Austin, and 4) the likelihood of officially dropping out of high school. Although we do not formally test mediation in this study, one possible mechanism by which reclassification may impact graduation is through the achievement outcomes discussed above. Because students must both pass the five EOC exams and earn credits in core subjects to receive their diplomas, positive impacts of reclassification on achievement may also improve graduation outcomes; negative impacts could create additional barriers to graduating.

Empirical Strategy

To estimate the effect of eighth-grade reclassification on students' subsequent high school outcomes, we use a regression discontinuity design that exploits Texas's use of test thresholds during the reclassification process. Because students who reclassify as fully English proficient are likely to differ systematically from their peers who continue as ML-EL, simple comparisons between their outcomes would not provide causal estimates of the treatment effect. However, students scoring within a narrow range around the test threshold are arguably similar on average across observable and unobservable characteristics (Imbens & Lemieux, 2008; Lee & Lemieux, 2010; Robinson, 2011).

Because Texas employs multiple reclassification criteria, we use a multiple-rating regression discontinuity design (Reardon & Robinson, 2012; Wong et al., 2013). First, we limit our sample to students who demonstrated English language proficiency in the domains of speaking, writing, and listening by scoring "Advanced High" on each corresponding section of TELPAS. We then use students' raw scores on the second administration of the eighth-grade STAAR Reading exam as our running variable that predicts reclassification. Put another way, if Texas did not include the state ELA content exam as part of the reclassification process, all students in our analytic sample would be eligible to reclassify based on TELPAS performance. Our frontier design allows us to explore the extent to which including an additional content exam criterion either creates an unnecessary barrier or adds a necessary check that ensures students do not lose access to EL services too early.

We first estimate the effect of attaining reclassification *eligibility* based on STAAR reading. This intent-to-treat (ITT) analysis uses the following general equation:

$$(1)\ Y_{it} = \beta_0 + \beta_1 Met_{it} + f(score_{it}) + \beta_2 Met_{it} * f(score_{it}) + \delta_t + X_{it} + \varepsilon_{it}$$

where *Y* is a generic outcome representing our outcomes of interest. *Met* is an indicator that equals one if student *i*, who was a ML-EL in eighth-grade cohort *t*, met the state policy threshold in STAAR Reading for reclassification. The running variable, *score*, is students' raw score on the second administration of STAAR Reading, recentered so that 0 is the passing score. f() is a flexible function linking test scores to outcomes. We include interactions to allow this relationship to vary on either side of the cutoff. Our coefficient of interest is β_1, which represents the effect of passing the reclassification threshold on students' subsequent outcomes. All models include cohort fixed effects (δ_t) to account for variations in how test scores were scaled across years. Our preferred specification uses local linear regression (Gelman & Imbens, 2019) and adds student covariates (X_{it}), including student demographics, achievement in seventh grade, and number of years enrolled in high school; eighth-grade school covariates, such as demographics and average achievement; and dummy variables for school region to increase precision.

Because the decision to reclassify a ML- EL student in Texas is based on committee (LPAC) deliberation and incorporates teacher evaluations, students who attain the policy thresholds may remain EL-classified, while LPAC discretion may allow others to be reclassified despite missing the cutoff. In addition to the ITT estimates from Equation 1, we employ a fuzzy RD design to estimate a "complier average treatment effect" (CATE), defined as students whose reclassification outcome was induced by their performance on the STAAR Reading test. We use the following two-stage least squares equations:

$$(2)\ Reclassified_{it} = \beta_0 + \beta_1 Met_{it} + f(score_{it}) + \beta_2 Met_{it} * f(score_{it}) + \delta_t + X_{it} + \varepsilon_{it}$$

$$(3)\ Y_{it} = \delta_0 + \delta_1 \widehat{reclassified}_{it} + f(score_{it}) + \delta_2 Met_{it} * f(score_{it}) + \delta_t + X_{it} + \varepsilon_{it}$$

The first stage regression (Equation 2) predicts the probability of reclassification. The second stage (Equation 3) uses the predicted probability of reclassification to estimate the outcome of interest for compliers.

All models for both the ITT and CATE are estimated using local polynomial non-parametric regression and the *rdrobust* package in Stata (Calonico et al., 2017). We use the optimal bandwidth from the first stage estimation (Equation 2) (BW=5 points) for all of our main results using (Calonico et al., 2014). We demonstrate the robustness of our results to alternative bandwidths below. Given that we expect a smooth transition in characteristics around the cutoff, we use a triangular kernel (Chin, 2021). Standard errors are heteroskedasticity-robust and clustered at the eighth-grade school level (Kolesár & Rothe, 2018).

Identifying Assumption Checks

To yield valid estimates, our RD approach assumes that potential outcomes are continuous across the treatment threshold. We conduct three analyses that suggest we are very likely to meet this assumption. First, we fit a series of models of the form specified in Equation 1, replacing outcome Y with student-level observable characteristics. We find no evidence of meaningful differences between students who do and do not meet the threshold in the area around the cutoff (Appendix A, Table A4).

Second, we explore the possibility of test manipulation. One concern is that if students can easily manipulate their score to be just above or below the cutoff for treatment, treated individuals may differ systematically from untreated individuals, particularly across unobserved characteristics, thus invalidating our RD design. In our context, score manipulation is unlikely because STAAR is graded by unaffiliated scorers at a centralized state testing center and students are very unlikely to predict the exact number of correct answers associated with scoring close to the cutoff. Nevertheless, we test for evidence of score manipulation in two ways. We first plot the density of raw STAAR Reading scores around the reclassification threshold to visually assess the possibility of manipulation (Appendix B, Figure B2). Table A5 (Appendix A) presents results for density tests for a smooth distribution of scores around the threshold and shows that we do not reject smoothness in any year (Cattaneo et al., 2018).

Our two-stage least squares instrumental-variables approach to estimate the CATE requires two additional identifying assumptions. First, we assume a strong first stage such that, conditional on meeting the TELPAS requirements for reclassification, passing the STAAR Reading threshold significantly increases the likelihood that students reclassify out of EL status. Figure 1 plots the probability of reclassification as a function of students' scores on the second administration of STAAR and shows that the estimated probability of eighth-grade reclassification increases by about 39 percentage points around the threshold.

[FIGURE 1]

Second, we assume that the exclusion restriction holds, meaning that the instrument—scoring above the running variable threshold—*only* affects our outcomes of interest through the reclassification process. One concern is that the STAAR Reading reclassification threshold is the cutoff for passing the exam itself. Failing STAAR reading multiple times is one condition for receiving accelerated instruction and grade retention, so our control group might have received extra reading support.[6] We did not identify any other significant interventions that students receive based on meeting the passing threshold for STAAR Reading, although we acknowledge that passing STAAR may be related to ninth-grade course placement in some schools. We discuss this possibility and potential effects of additional reading support in greater detail below.

Results

Impacts on High School Course Taking

For students near the eighth-grade STAAR reading threshold who were relatively proficient in English, our results suggest that reclassifying at the end of eighth grade increased the likelihood of earning Advanced Placement, International Baccalaureate, or Dual Enrollment course credit (hereafter “advanced”) in high school. Table 2 presents ITT (Columns 1 and 2) and CATE (Columns 3 and 4) estimates for course taking outcomes from our main RD models, using the optimal bandwidth from the first stage. All models include year fixed effects. Columns 2 and 4—our preferred specifications— add pretreatment student, eighth-grade school, and district region covariates to increase precision. In this context, the ITT estimates capture the expected overall effect of including academic criteria in the state guidelines for reclassification given imperfect reclassification rates for students who meet the academic thresholds, whereas the CATE estimates capture the expected impact on outcomes for students who actually reclassify based on meeting academic performance thresholds. We argue that both estimands are policy relevant at different organizational levels—namely, the state versus individual levels.

[TABLE 2]

We estimate that attaining reclassification eligibility by scoring just above the cutoff increased the likelihood of earning credit in any advanced course by roughly 3 percentage points (10% of students below the threshold in the “control” condition earn advanced credit at baseline) and the likelihood of earning multiple advanced credits by 2 percentage points (5% below threshold). Actually exiting EL status increased the likelihood of any advanced credit attainment by 8 percentage points. These gains seem to be concentrated in the humanities. Our ITT models predict 2 percentage point increases in the likelihood of earning advanced English and 3 percentage point increase in the likelihood of earning advanced social studies credit (4% and 7% below threshold). CATE estimates increase to 5 percentage points in English and 6 percentage points in social studies, nearly doubling the likelihood of earning credits in these advanced courses for ML-ELs who reclassify before high school matriculation, relative to their peers who enter under continued EL status. Our point estimates for the change in likelihood of earning advanced credit in math and science are positive across models but not significant at conventional levels.

Impacts on High School Achievement

Tables 3 and 4 present estimates for our achievement outcomes, including performance on the first attempt of the five end-of-course exams required for high school graduation in Texas and the likelihood of failing core courses in high school. Among students close to the threshold, just attaining reclassification eligibility resulted in a negative effect on initial Algebra I EOC scores, roughly 0.05 SD in magnitude (Table 3). Reclassifying out of EL status based on this eligibility resulted in a negative effect 0.13 SD in magnitude. Given the distribution of Algebra I EOC scores among ML-ELs, this decrease indicates that reclassified students scored at approximately the 40th percentile while their peers who remained in the EL-status scored at the 33rd percentile. Additionally, we find suggestive evidence that reclassified students are more likely to retake the Algebra I EOC, which could indicate that score declines increase the likelihood of failing the exam. Our ITT estimates suggest that students just above the threshold are 3 percentage points more likely to retake the Algebra I exam at least once (relative to 29% below the threshold). Students who actually reclassify are 7 percentage points more likely to retake. These results are marginally significant under our preferred specifications in columns 2 and 4. Estimates for other exam scores are generally negative, apart from biology, but are small in magnitude and insignificant.

[TABLE 3]

We also find suggestive evidence of negative impacts on course performance (Table 4). Our ITT results indicate that students just above the threshold are approximately 3 to 4 percentage points more likely to fail at least one course in English (relative to 55% below threshold), science (50% below threshold), and social studies (47% below threshold). CATE results in Column 4 suggest that reclassified students are 11 percentage points more likely to fail a class in English, 10 percentage points more likely to fail in science, and 8 percentage points more likely to fail in social studies. Although these results are only marginally significant at conventional levels, they suggest that among students with otherwise similar English proficiency and academic profiles, reclassification may lead to increased academic struggle, both on academic assessments and within their core courses.

[TABLE 4]

Impacts on High School Graduation

Finally, we find evidence that reclassifying at the end of eighth grade negatively impacts a range of graduation outcomes (Table 5). Students who score just above the threshold are 4 percentage points less likely to graduate on time, defined as in the spring of the fourth year after entering high school (83% below threshold); 3 percentage points less likely to graduate “college-ready,” i.e., with the coursework required for a selective college (11% below threshold); and 2 percentage points more likely to drop out of high school entirely (9% below threshold) (Column 2). Reclassified students are 5 percentage points less likely to ever graduate, 10 percentage points less likely to graduate on time, 7 percentage points less likely to graduate with college preparatory coursework, and 5 percentage points more likely to drop out (Column 4). All results are statistically significant at conventional levels and stable across models.

[TABLE 5]

Robustness Checks

We verify the robustness of our results in several ways. First, in our main results in Tables 2-5, we show that results are robust to the inclusion of student, school, and district-region covariates. Although the magnitudes of some estimates differ by more than 10% with the inclusion of covariates, the direction and practical significance of all estimates remain the same across outcomes. In Appendix A Tables A6-A9, we demonstrate that results are also robust to specifying multiple different bandwidths from 4 to 10 points around the threshold (Chin, 2021; Robinson-Cimpian & Thompson, 2016). Specifying bandwidths is a trade-off between increasing precision as the bandwidth expands and reducing bias as the bandwidth shrinks (Brunner et al., 2023). Our estimates generally increase in magnitude when we apply a smaller bandwidth. In Tables A10-A13, we run a series of falsification tests in which we shift the threshold for reclassification in both directions by 4 points. Point estimates are always statistically insignificant and standard errors are large. We also show that our results are robust to excluding the 2016 and 2017 cohorts, who completed high school during the COVID-19 pandemic (Tables A14-A17, Appendix A).

Finally, to mitigate concerns about attrition we conducted two tests. First, we test for differential attrition among students scoring near the cutoff. Specifically, we estimate Equation 1 where the dependent variable takes the value of 1 if the student has information on the outcome of interest and 0 otherwise. Appendix A Tables A18-21 show the results of this exercise for each outcome. We do not find evidence of differential attrition across outcomes. Our second check for attrition consisted of estimating all outcomes for the same analytic sample. When we condition the sample for students who have information on all outcomes, the effects of reclassification on course taking and graduation outcomes are very consistent with our main specification (See Tables A22-25, Appendix A). The negative effect of reclassification on Algebra I standardized scores is no longer statistically significant. Overall, our attrition analyses suggest that our main results are not driven by students leaving the sample.

Discussion

This study provides new causal evidence on two aspects of the EL reclassification process—timing and determining criteria—using statewide data from Texas, a highly salient policy context that serves the highest percentage of ML-ELs and the second largest population in absolute numbers (TEA, 2024). Overall, our results suggest that reclassifying at the end of eighth grade and entering high school without the EL label can be both a bulwark and a barrier to subsequent outcomes for students. We find that for the marginal student in our sample, entering high school as non-EL positively impacts their course taking outcomes, but negatively impacts achievement and graduation.

Our finding that reclassifying in eighth grade increases the likelihood of earning credit in advanced courses in high school provides some of the first causal evidence that EL classification limits students' access to rigorous coursework (Callahan et al., 2010; Callahan, 2005; Callahan & Shifrer, 2016; Johnson, 2019). Advanced Placement, International Baccalaureate, and dual enrollment courses can confer numerous benefits, including access to more rigorous instruction, the opportunity to earn college credits, and advantages in selective admissions processes (Warne, 2017). Our estimates suggest that removing the STAAR Reading reclassification criteria would increase the likelihood of successfully completing an advanced course by about three percentage points for the marginal student, a modest but important change given that only 6% of eighth-grade ML-ELs ever earn credit in these courses. Because AP, IB and DE classes are usually taken in later grades, our results also suggest that eighth-grade reclassification may have long-reaching implications for students' high school trajectories.

Simultaneously, our results suggest that many of the students who reclassified before entering high school by meeting the STAAR Reading threshold struggled academically in their core classes. They were more likely to fail courses in multiple subjects and had lower standardized test scores in Algebra I. Although some students both pass advanced courses and fail core courses, most of the students who fail core content courses did not complete advanced coursework. Our combined course failure and testing results suggest that on average, Texas students who reclassify by just barely meeting the eighth-grade STAAR Reading threshold may be struggling more across *all four* core content areas in high school, which is alarming.[7] Prior research rarely examines achievement outcomes aside from test scores, making it difficult to contextualize our results within the existing literature. However, one study conducted by Pope (2016) in California's Los Angeles Unified School District found that reclassification in middle and high school had no effect on subsequent math and ELA GPA; the apparent divergence of our findings from Pope's emphasizes the highly contextual nature of reclassification effects.

Appendix A Tables A27-30 provide some suggestive evidence that access to EL services in academic classes may be a driving mechanism behind some of the observed negative impacts of reclassifying as "Fully English Proficient." As discussed above, most secondary ELs in Texas are either placed in ESL for all content courses or ESL for ELA only (Table A1). We run two separate analyses, one using a subsample of schools where ESL All Content is the primary instructional program for ML-ELs and one using a subsample of schools with ESL ELA is the primary program. These results provide some suggestive evidence that the impacts of reclassification are more consistently negative in schools that offer ESL in all content courses. For example, we estimate that compared to peers who were likely placed in an ESL math class, reclassified students score about 0.2 standard deviations lower on their Algebra I EOC test (not significant at conventional levels). Our point estimate for schools with ESL in ELA Only is much smaller. Similarly, point estimates for the increased likelihood of core course failure and decreased likelihood of graduating on-time or college ready are generally larger in ESL all content schools. These patterns could indicate the importance of ESL services across academic content areas, even for relatively high English proficiency students like those in our sample.

Prior research suggests that schools with larger ML-EL populations tend to offer more comprehensive and effective services for students who are learning English (Umansky & Reardon, 2014). Tables A31-34 provide additional evidence that negative results of reclassification may be driven by losing access to effective EL support services. We run two separate analyses: one for the subgroup of students who attend schools in the bottom third in terms of ML-EL population (0-8%) and one for students who attend schools in the top third (15%+). Estimates are imprecise due to small sample size. Overall, however, our results suggest that the impacts of reclassification on test scores and long-term outcomes like graduation are more consistently large and negative in schools with larger ML-EL populations. If schools that serve more ML-ELs have more comprehensive supports, these findings may suggest that observed negative impacts of reclassification are driven by students losing access to effective services.

Finally, evidence that students who exit EL status prior to high school are 4 percentage points less likely to graduate on time than their peers who remained ML-EL is concerning, given the importance of a high school diploma for economic and social well-being in the US (Campbell, 2015). Our achievement results offer one potential mechanism for this finding, as passing EOC exams and accruing core subject credits are graduation requirements in Texas. In Table A35 (Appendix A), we present results from an exploratory analysis showing that reclassified students were less likely to accrue the total number of credits in English and science needed to graduate. These results, along with suggestive evidence from our course taking results, also help explain why reclassified students were less likely to graduate "college-ready," with the appropriate number and type of credits in each core content area to gain admission to many four-year institutions in Texas. Reclassified students in our sample appeared somewhat more likely to take remedial math courses, which could divert them from the college-prep math track, although our coefficients did not reach statistical significance.

Taken together, our achievement and graduation findings suggest that high English proficiency, lower academic proficiency students in our sample who reclassify in eighth grade continue to struggle academically on average—and actually perform worse than similar peers who matriculate as EL-classified, with meaningful consequences for high school completion. The negative effects of reclassification seem to be most pronounced in schools with more comprehensive EL services and larger populations of ML-ELs.

Generalizability and Limitations

Our regression discontinuity design offers strong internal validity. However, our sample restrictions influence the external validity of our results. Specifically, our results apply to a narrow but important subgroup: students with relatively high English proficiency who score near the reclassification threshold on their second STAAR Reading attempt. We argue that this group remains highly policy relevant for several reasons. First, the combination of strong oral proficiency and weaker academic literacy is common among secondary ELs and particularly among long-term ML-ELs (Fu, 2021). Second, while ELP exam retakes are less common, many states offer opportunities to retake the state ELA content exam. All three of the states that currently mandate ELA exams for reclassification allow students to retake the exam in some grades. Thus, our study sheds light on a scenario that parallels existing policies outside Texas.

As discussed above, two other factors distinguish Texas' use of ELA exams during our study period. First, Texas used ELA as a substitute for ELP reading, not as an additional criterion. Our results do not speak directly to the impacts of requiring both ELA and ELP, which is now the criterion used in Texas, as well as New York and Florida. However, other states like Nevada and South Dakota currently allow students who do not meet ELP requirements to reclassify based on ELA performance. Further, given the frequent shifts in reclassification policies over time (Morales & Lepper, 2024), our study offers valuable evidence about the effects of using an ELA test as the determining reading criteria for reclassification. The passing threshold for STAAR ELA in Texas was also relatively low compared to the TELPAS ELP threshold. Indeed, more than half of all ML-ELs in our sample passed the ELA content exam each year. We can think of our results as providing insight into two policy choices: 1) emphasizing critical reading skills by using the ELA test instead of the ELP test; 2) holding students to a relatively low standard for demonstrating reading proficiency. In general, we find that using only an ELA test with a lower passing threshold allowed some students to lose access to support before they were ready.

Implications for Policy, Practice, and Research

Despite their locality, our findings have substantive implications for ML-EL secondary education. Below, we highlight three major areas for action from policymakers and practitioners at multiple organizational levels—state, district, and school—which can be facilitated by further scholarly advancements.

Increasing ML-ELs' Access to Advanced Coursework

In spite of federal guidance stipulating that EL-classified students are entitled to equal access to AP, IB, and honors coursework (U.S. Department of Education, Office for Civil Rights, 2015), a large body of descriptive research suggests that ML-ELs have constrained access to advanced and college-preparatory classes, even after accounting for factors like prior achievement (Callahan et al., 2010; Callahan, 2005; Callahan & Shifrer, 2016; Kanno & Kangas, 2014). Our RD results provide causal evidence that EL status itself acts as a barrier to taking advanced courses, even for students with relatively high English language proficiency. To help mitigate these effects, future research should explore the characteristics—e.g., test performance, course grades, interest, access to school resources—that predict success in advanced classes for students who are ML-ELs in middle school. School leaders can then draw on those characteristics to proactively recruit multilingual students to enroll in advanced classes, regardless of EL status (Ricciardi & Winsler, 2021).

Schools should also reevaluate practices that create barriers to AP participation for students from marginalized backgrounds, such as requiring that students receive a teacher recommendation or that they take the AP exam if they enroll in the class without providing a fee voucher (S. Flores & Gomez, 2011). Ensuring that instruction in advanced courses is accessible for ML-ELs is a crucial part of meaningfully expanding access to advanced content (Kanno & Kangas, 2014). Finally, future research should explore the extent to which the supply of AP and other advanced courses is diminished at the schools that ML-ELs typically attend. If studies uncover significant disparities, district and state leaders can take steps to continually monitor school course offerings and potentially redistribute resources to guarantee more equitable access for all students.

Reassessing the Use of Academic Content Criteria in Reclassification Decisions

This study also provides new evidence on the appropriateness of requiring students to demonstrate ELA content proficiency in order to reclassify, a practice used by several states that serve the largest numbers of ML-ELs. Our results somewhat support arguments that using ELA to assess reading proficiency may decrease access to advanced courses among some students who are ready to succeed in these classes, particularly in English. We find that reclassified students are more likely to earn credit in AP, IB, or dual enrollment English classes. However, our results indicate that even among students who have met all three other test-based criteria—including the high bar of demonstrating “advanced high” proficiency on the TELPAS writing ELP exam—reclassification reduces overall achievement on average. Reclassifiers are more likely to fail courses, score lower on a key graduation exam, and are less likely to graduate high school on time. Contrary to the argument that including content criteria may be particularly harmful for long term ML-ELs, our results suggest that for a specific subgroup of students who demonstrate strong English fluency but struggle with critical reading skills academic content, simply requiring them to meet the minimum passing standard for ELA may be insufficient to ensure they are ready to succeed in “mainstream” instructional settings without additional services.

These results could indicate that the criteria for reclassification should be made more stringent, a step that Texas has already taken by adding the ELP reading exam as a reclassification criterion. However, given the policy context in Texas, it is difficult to disentangle if negative effects of reclassification based on ELA only are driven by the content assessed on the ELA exam or by the lower passing standard applied. Future research should explicitly test the effects of requiring students to pass both ELP and ELA tests to offer further insights into the implications of different ways that states choose to incorporate ELA into the reclassification process.

Supporting Multilingual Learners' Academic Success and High School Completion

Our findings that reclassified students struggle academically and are subsequently less likely to graduate than peers who enter high school with similar English proficiency and academic skills are troubling and invite policy action. Although we explore a few potential explanations for this phenomenon, our evidence on the precise mechanisms linking EL classification to achievement for the marginal student remains suggestive, necessitating further research. We find some evidence that negative impacts on achievement may be driven by decreased access to beneficial EL services, particularly to EL-specific versions of content courses outside of ELA. If future research can identify particular services that benefit students who are nearing reclassification, continuing to provide these services during the transition out of EL status could be a good strategy to improve outcomes. For Texas, specifically, researchers should also extend prior work on how LPACs make decisions about testing accommodations for ML-ELs in the face of accountability pressures (Mavrogordato & White, 2020), and investigate how the sudden loss of such accommodations in non-ELA content areas might influence achievement for reclassified students. Policymakers might then consider whether accommodations in content areas like math should be allowable for students in the first few years of post-exit monitoring and adjust state guidelines accordingly.

Conclusion

Policy and programming for multilingual learners should enable their educational thriving; classification and reclassification are a few means to that end. Our study highlights the need for more holistic, nuanced approaches in research, policy, and practice that attend to the complexity of multidimensional reclassification effects. When ML-EL students reclassify, how they reclassify, who reclassifies based on the criteria instituted, and what changes when they reclassify are all important to understand in determining student placement and the subsequent support that they receive. Overall, our findings provide strong evidence that academic-based reclassification in eighth grade at the critical middle-to high school transition can be extremely consequential for ML-ELs' high school experiences and outcomes. Attention to how the transition out of EL status affects different dimensions of the educational environment can guide more targeted policy and practice changes to support students during, and after, the process to improve their odds of long-term success.

End Notes:

1. We follow Bartlett et al. (2024) in using the term "ML-EL" as an assets-based way to refer to the subset of multilingual students who are officially identified for EL services. When referring to the services or policies themselves we use "EL" as the descriptor to align with the language of federal guidelines (e.g., "EL programming," "EL status").

2. Our analysis assumes that all districts used TELPAS as their oral and writing proficiency measures for reclassification during our study period, although state policy permitted districts to use other approved ELP assessments. Evidence from our first stage estimation suggests that this is a reasonable assumption.

3. We identify student program placement using program flags and definitions from Texas' longitudinal data system. The two ESL program types, ESL Content and ESL Pull-Out/ELA, are defined in state legislation. TEA websites also include extensive resources and toolkits around implementing each model. However, the specific implementation of each program likely varies widely across schools and districts.

4. There were up to four administrations of STAAR in a year. Eleven percent of students who took STAAR on the third administration met the standard and 12% who took the fourth administration met the standard. Most students took either the third or fourth administration but not both. Only 4% of students took both administrations. Because so few students took the third or four administrations, we focus on the second administration to create a relatively clean comparison between those who did and did not qualify to reclassify.

5. Using scale scores leads to mass points being detected in the running variable and few unique observations within the optimal bandwidth upon which to estimate reclassification effects. For instance, in 2015, the passing scale score was 1575 (raw score of 27). The gap between passing and not passing is 9 scale points. In 2017, however, the passing scale score was 1587 (raw score of 24). There is no possibility of scoring 1575 in 2017 and the gap between 1576, the last non-passing score, and the cut score is 11 points. For more consistent bandwidths in the RD, we chose to use standardized raw scores. We obtain passing thresholds from conversation tables which are publicly available on the Texas Education Agency website.

6. According to the Texas Education Code, students who failed STAAR for the third time should receive accelerated instruction in the following year at least once a week, in small classes and with a certified teacher. However, each school district decided when the program would be delivered (i.e., during school hours or after school hours operations), how many times per week, and what curriculum to follow.

7. It may seem somewhat paradoxical that our results suggest reclassified students are both more likely to earn credit in advanced English and social studies courses and more likely to fail classes in the same subjects. One potential explanation for this phenomenon is that effects on each outcome are driven by different student subgroups above the threshold. Most of the reclassified students who take advanced social studies and English do not fail their core courses in these subjects (about 70% of advanced course takers in the sample do not fail a core humanities course). Put another way, the group of reclassified students who are failing core humanities courses are largely a different subgroup than those who are passing advanced courses. In Appendix A, Tables A27 and A31, we run heterogeneity analyses to assess whether advanced course and course failure results are driven by students at different types of schools. We do not find strong evidence that results are concentrated in schools with a certain level of ML-EL population or type of EL program.

References

Abedi, J. (2006). Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record*, *108*(11), 2282–2303. https://doi.org/10.1111/j.1467-9620.2006.00782.x

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, *27*(3), 17–31. https://doi.org/10.1111/j.1745-3992.2008.00125.x

Bailey, A. L., & Carroll, P. E. (2015). Assessment of English language learners in the era of new academic content standards. *Review of Research in Education*, *39*, 253–294.

Baker, S., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., Gersten, R., Haymond, K., Kieffer, M. J., Linan-Thompson, S., & Newman-Gonchar, R. (2014). Teaching academic content and literacy to English learners in elementary and middle school. IES practice guide. NCEE 2014-4012. In *What Works Clearinghouse*. What Works Clearinghouse. https://eric.ed.gov/?id=ED544783

Bartlett, C., Callahan, R., & Mavrogordato, M. (2024). Beyond a transformative approach and deep understanding: External factors and mid-level leaders' policy implementation to expand equity for English learners. *Educational Administration Quarterly*, *60*(2), 151–190. https://doi.org/10.1177/0013161X241230296

Betts, J., Hill, L., Bachofer, K., Hayes, J., Lee, A., & Zau, A. (2020). *Effects of English learner reclassification policies on academic trajectories* (No. w28188). National Bureau of Economic Research. https://doi.org/10.3386/w28188

Brunner, E. J., Dougherty, S. M., & Ross, S. L. (2023). The effects of career and technical education: Evidence from the Connecticut technical high school system. *The Review of Economics and Statistics*, *105*(4), 867–882. https://doi.org/10.1162/rest_a_01098

Callahan, R. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, *42*(2), 305–328.

Callahan, R., & Shifrer, D. (2016). Equitable access for secondary English learner students: Course taking as evidence of EL program effectiveness. *Educational Administration Quarterly*, *52*(3), 463–496. https://doi.org/10.1177/0013161X16648190

Callahan, Wilkinson, L., & Muller, C. (2010). Academic achievement and course taking among language minority youth in U.S. Schools: effects of ESL placement. *Educational Evaluation and Policy Analysis*, *32*(1), 84–117. https://doi.org/10.3102/0162373709359805

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(2), 372–404. https://doi.org/10.1177/1536867X1701700208

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs: Robust nonparametric confidence intervals. *Econometrica*, *82*(6), 2295–2326. https://doi.org/10.3982/ECTA11757

Campbell, C. (2015). The socioeconomic consequences of dropping out of high school: Evidence from an analysis of siblings. *Social Science Research*, *51*, 108–118. https://doi.org/10.1016/j.ssresearch.2014.12.011

Carlson, D., & Knowles, J. E. (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. https://doi.org/10.1002/pam.21908

Cashiola, L., Cigarroa Kennedy, C., Molina, M., Ma, H., & Varghese, L. (2022). *Long-term English learners: How is timing of reclassification associated with middle and high school outcomes?* Rice University Kinder Institute for Urban Research. https://doi.org/10.25611/WGNN-Z360

Cattaneo, M. D., Jansson, M., & Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, *18*(1), 234–261. https://doi.org/10.1177/1536867X1801800115

Chin, M. J. (2021). The effect of English learner reclassification on student achievement and noncognitive outcomes. *Journal of Research on Educational Effectiveness*, *14*(1), 57–89. https://doi.org/10.1080/19345747.2020.1831116

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English learner reclassification policy effects across districts. *American Educational Research Journal*, *54*(1_suppl), 255S-278S. 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 learner classrooms. *Journal of Education for Students Placed at Risk (JESPAR)*, *19*(2), 98–124. https://doi.org/10.1080/10824669.2014.954044

Ensuring English learner students can participate meaningfully and equally in educational programs. (2015). US Department of Education.

Estrada, P. (2014). English learner curricular streams in four middle schools: Triage in the trenches. *The Urban Review*, *46*(4), 535–573. https://doi.org/10.1007/s11256-014-0276-7

Flores, N., Kleyn, T., & Menken, K. (2015). Looking holistically in a climate of partiality: identities of students labeled long-term English language learners. *Journal of Language, Identity & Education*, *14*(2), 113–132. https://doi.org/10.1080/15348458.2015.1019787

Flores, S., & Gomez, M. O. (2011). Strategies for increasing advanced placement participation for underrepresented students: Barriers, practices, and positive outcomes. *NASSP Bulletin*, *95*(1), 65–79. https://doi.org/10.1177/0192636511406529

Forte, E., Perie, M., & Paek, P. (2012). *Exploring the relationships between english language proficiency assessments and english language arts assessments*. Ed Count.

Fu, S. (2021). Breaking the silence: A critical review of language policy and planning for long-term English learners. *GATESOL Journal*, *31*(1), 17–35. https://doi.org/10.52242/gatesol.125

Gamoran, A. (2009). *Tracking and inequality: New directions for research and practice.* Wisconsin Center for Education Research, 6. http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Gándara, P. (2021). The gentrification of two-way dual language programs: A commentary. *Language Policy*, *20*(3), 525–530. https://doi.org/10.1007/s10993-021-09595-z

Gándara, P., & Baca, G. (2008). NCLB and California's English language learners: The perfect storm. *Language Policy*, *7*(3), 201–216. https://doi.org/10.1007/s10993-008-9097-4

Gándara, P., Rumberger, R., Maxwell-Jolly, J., & Callahan, R. (2003). English learners in California schools: Unequal resources, 'unequal outcomes. *Education Policy Analysis Archives*, *11*, 36-. https://doi.org/10.14507/epaa.v11n36.2003

Garibay & Abdelkader. (2024). *Office of English Language Acquisition: Reclassifying English Learner Students*. https://ncela.ed.gov/sites/default/files/2024-08/reclassifyingels-20240806-508.pdf

Gifford, B. R., & Valdés, G. (2006). Chapter 7 The linguistic isolation of Hispanic students in California's public schools: The challenge of reintegration. *Yearbook of the National Society for the Study of Education*, *105*(2), 125–154. https://doi.org/10.1111/j.1744-7984.2006.00079.x

Hopkins, M., Weddle, H., Lavadenz, M., Murillo, M. A., & Vahdani, T. (2022). Examining the English learner policy ecology: How educators navigated the provision of designated English language development (ELD) support at the secondary level. *Peabody Journal of Education*, *97*(1), 47–61. https://doi.org/10.1080/0161956X.2022.2026719

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Mann, F. B., Barmer, A., Dilig, R., Nachazel, T., Barnett, M., & Purcell, S. (2020). *The Condition of Education 2020*. 348.

Hwang, N., & Koedel, C. (2023). Helping or hurting: The effects of retention in the third grade on student outcomes. *Educational Evaluation and Policy Analysis*, 01623737231197639. https://doi.org/10.3102/01623737231197639

Johnson, A. (2019). The effects of English learner classification on high school graduation and college attendance. *AERA Open*, *5*(2), 2332858419850801. https://doi.org/10.1177/2332858419850801

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/10.3102/0162373719877197

Kanno, Y., & Kangas, S. E. N. (2014). '"I'm not going to be, like, for the AP"': English language learners' limited access to advanced college-preparatory courses in high school. *American Education Research Journal*.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304. https://doi.org/10.1257/aer.20160945

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. https://doi.org/10.3102/01623737221097419

Lillie, K. E., Markos, A., Arias, M. B., & Wiley, T. G. (2012). Separate and not equal: The implementation of structured English immersion in Arizona's classrooms. *Teachers College Record*, *114*(9), 1–33. https://doi.org/10.1177/016146811211400906

Linquanti, R. (2001). *The redesignation dilemma: Challenges and choices in fostering meaningful accountability for English learners*. University of California Linguistic Minority Research Institute.

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis*, *36*(4), 457–475. https://doi.org/10.3102/0162373714527788

Long, M., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, *49*(2). https://journals-sagepub-com.proxy.library.vanderbilt.edu/doi/10.3102/0002831211431952

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different skills: Identifying differentially effective teachers of English language learners. *The Elementary School Journal*, *117*(2), 261–284.

Mavrogordato, M., & White, R. S. (2020). Leveraging policy implementation for social justice: How school leaders shape educational opportunity when implementing policy for English learners. *Educational Administration Quarterly*, *56*(1), 3–45. https://doi.org/10.1177/0013161X18821364

Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on "long-term English language learners": Characteristics and prior schooling experiences of an invisible population. *International Multilingual Research Journal*, *6*(2), 121–142. https://doi.org/10.1080/19313152.2012.665822

Morales, & Lepper. (2024). *The changing landscape of states' English learner reclassification policies.* Center for Education, Efficacy, Excellence and Equity (E4). https://e4.northwestern.edu/2024/04/25/the-changing-landscape-of-states-english-learner-reclassification-policies/

Olsen. (2010). *Reparable harm: Fulfilling the unkept promise of educational opportunity for California's long term English learners*. https://californianstogether.org/reparable-harm-fulfilling-the-unkept-promise-of-educational-opportunity-for-californias-long-term-english-learners/

Pope, N. G. (2016). The marginal effect of K-12 English language development programs: Evidence from Los Angeles Schools. *Economics of Education Review*, *53*, 311–328. https://doi.org/10.1016/j.econedurev.2016.04.009

Porter, Vazquez Cano, & Umanksy. (2023). *Bilingual education and America's future: Evidence and pathways*. The Civil Rights Project.

Rafa, A., Erwin, B., Brixey, E., McCann, M., & Perez Jr., Z. (2020). *50-state comparison: English learner policies*. Education Commission of the States. https://reports.ecs.org/comparisons/50-state-comparison-english-learner-policies-13

Ragan, A., & Lesaux, N. (2006). Federal, state, and district level English language learner program entry and exit requirements: Effects on the education of language minority learners. *Education Policy Analysis Archives*, *14*, 20-. https://doi.org/10.14507/epaa.v14n20.2006

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. https://doi.org/10.1080/19345747.2011.609583

Reyes, M., & Hwang, N. (2021). Middle school language classification effects on high school achievement and behavioral outcomes. *Educational Policy*, *35*(4), 590–620. https://doi.org/10.1177/0895904818823747

Ricciardi, C., & Winsler, A. (2021). Selection into advanced courses in middle and high school among low-income, ethnically diverse youth. *Journal of Advanced Academics*, *32*(3), 291–323. https://doi.org/10.1177/1932202X21990096

Robinson, J. P. (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/10.3102/0162373711407912

Robinson-Cimpian, J. P., & Thompson, K. D. (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/10.1002/pam.21882

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher*, *37*(4), 189–199. https://doi.org/10.3102/0013189X08319569

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. https://doi.org/10.3102/0002831216634463

Thompson, K. D. (2017). What blocks the gate? Exploring current and former English learners' math course-taking in secondary school. *American Educational Research Journal*, *54*(4), 757–798. https://doi.org/10.3102/0002831217706687

Torre Gibney, D., & Henry, G. (2020). Who teaches English learners? A study of the quality, experience, and credentials of teachers of English learners in a new immigrant destination. *Teaching and Teacher Education*, *90*, 102967. https://doi.org/10.1016/j.tate.2019.102967

Umansky. (2016a). Leveled and exclusionary tracking: English learners' access to academic content in middle school. *American Education Research Journal*. https://doi.org/10.3102/0002831216675404

Umansky. (2016b). To be or not to be EL: An examination of the impact of classifying students as English learners. *Educational Evaluation and Policy Analysis*, *38*(4), 714–737. https://doi.org/10.3102/0162373716664802

Umansky, & Avelar, J. (2023). Canaried in the coal mine: What the experiences and outcomes of students considered long-term English learners teach us about pitfalls in English learner education…and what we can do about it. *Journal of Education for Students Placed at Risk (JESPAR)*, *28*(1), 122–147. https://doi.org/10.1080/10824669.2022.2123326

Umansky, & Dumont, H. (2021). English learner labeling: How English learner classification in kindergarten shapes teacher perceptions of student skills and the moderating role of bilingual instructional settings. *American Educational Research Journal*, *58*(5), 993–1031. https://doi.org/10.3102/0002831221997571

Umansky, I. M., & Porter, L. (2020). State English learner education policy: A conceptual framework to guide comprehensive policy action. *Education Policy Analysis Archives*, *28*, 17. https://doi.org/10.14507/epaa.28.4594

Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*. https://doi.org/10.3102/0002831214545110

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, *14*(3–4), 139–159. https://doi.org/10.1080/10627190903425883

Wong, V. C., Steiner, P. M., & Cook, T. D. (2013). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*, *38*(2), 107–141. https://doi.org/10.3102/1076998611432172

Zuniga, K., Olson, J. K., & Winter, M. (2005). Science education for rural Latino/a students: Course placement and success in science. *Journal of Research in Science Teaching*, *42*(4), 376–402. https://doi.org/10.1002/tea.20064

TABLE 1. *Descriptive Characteristics of the Sample, 2013-2017*

	(1) All Grade 8 ML-ELs	(2) Analytic RD Sample (within 2 SD)
Male	55%	51%
Economically disadvantaged	88%	89%
Home language: Spanish	91%	95%
Race/Ethnicity		
Asian	4%	2%
Black	1%	1%
Hispanic/Latino	92%	95%
Multiracial	<1%	<1%
Native American	<1%	<1%
NHPI	<1%	<1%
White	2%	1%
Years in US Schools		
1-2 years	9%	1%
3-4 years	7%	3%
5+ years	72%	95%
SPED in Grade 8	12%	7%
Age	13.34	13.32
Observations	208,060	21,044
Met STAAR Standard		
Grade 7 Reading	31%	16%
Grade 7 Math	39%	33%
Grade 7 Writing	26%	20%
Observations	184,274	20,635

Note. Descriptive statistics for students enrolled in eighth grade in Texas public schools during the 2012-2013 to 2016-2017 academic years. Column 1 shows all English-learner classified eighth-grade students. Column 2 shows the analytic sample, which includes all ML-EL students who 1) have valid test scores on both TELPAS and STAAR reading; 2) scored ‘advanced high’ on the writing, speaking, and listening domains of TELPAS (requirements for reclassification); 3) failed the first STAAR attempt and re-took the second administration without receiving linguistic accommodations, 4) enrolled in a public school in Texas in the following grade; and 4) scored within 2 standard deviations around the passing cutoff. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness; SPED = Identified for Special Education.

TABLE 2. *Effects of Reclassification on Probability of Earning Advanced and Remedial Course Credit*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
AP/IB/DE courses				
Any course	0.0349***	0.0308**	0.0847***	0.0750**
	(0.0121)	(0.0121)	(0.0300)	(0.0297)
Control mean	[0.1008]	[0.1008]	[0.1008]	[0.1008]
>1 course	0.0220**	0.0198**	0.0534**	0.0482**
	(0.0088)	(0.0089)	(0.0216)	(0.0217)
Control mean	[0.0506]	[0.0506]	[0.0506]	[0.0506]
English	0.0247***	0.0222***	0.0600***	0.0539***
	(0.0080)	(0.0081)	(0.0197)	(0.0199)
Control mean	[0.0418]	[0.0418]	[0.0418]	[0.0418]
Math	0.0049	0.0039	0.0119	0.0096
	(0.0047)	(0.0048)	(0.0115)	(0.0116)
Control mean	[0.0137]	[0.0137]	[0.0137]	[0.0137]
Science	0.0030	0.0009	0.0074	0.0021
	(0.0060)	(0.0061)	(0.0146)	(0.0149)
Control mean	[0.0189]	[0.0189]	[0.0189]	[0.0189]
Social Studies	0.0263***	0.0257**	0.0640***	0.0624**
	(0.0100)	(0.0102)	(0.0247)	(0.0250)
Control mean	[0.0690]	[0.0690]	[0.0690]	[0.0690]
Remedial courses				
Math	0.0152	0.0206	0.0369	0.0500
	(0.0157)	(0.0158)	(0.0384)	(0.0386)
Control mean	[0.2020]	[0.2020]	[0.2020]	[0.2020]
Science	-0.0014	-0.0049	-0.0034	-0.0120
	(0.0206)	(0.0196)	(0.0501)	(0.0476)
Control mean	[0.3150]	[0.3150]	[0.3150]	[0.3150]
Observations	13,391	12,564	13,391	12,564

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.36 points), calculated via Calonico et al.'s (2014) bandwidth selection procedure. Robust standard errors, clustered at the eighth-grade school level, in parentheses. Control refers to students who scored below the STAAR cut score. All models include year fixed effects. Cols. 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region. ITT = Intent-to-treat. CATE = Complier local average treatment effect. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: "Math Models" and "Math Strategies." Remedial science = "Integrated Physics and Chemistry."
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 3. *Effects of Reclassification on High School Student Achievement*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
STAAR End-of-Course (EOC) Scores in SDs				
Algebra I	-0.0539**	-0.0529**	-0.1303**	-0.1284**
	(0.0267)	(0.0264)	(0.0653)	(0.0648)
Control mean	[-0.5239]	[-0.5239]	[-0.5239]	[-0.5239]
Observations	12,794	12,163	12,794	12,163
Biology	-0.0006	0.0108	-0.0014	0.0264
	(0.0262)	(0.0265)	(0.0635)	(0.0644)
Control mean	[-0.7523]	[-0.7523]	[-0.7523]	[-0.7523]
Observations	12,685	12,069	12,685	12,069
English I	-0.0127	-0.0158	-0.0308	-0.0385
	(0.0190)	(0.0190)	(0.0460)	(0.0462)
Control mean	[-0.6224]	[-0.6224]	[-0.6224]	[-0.6224]
Observations	12,829	12,195	12,829	12,195
English II	-0.0272	-0.0232	-0.0660	-0.0561
	(0.0245)	(0.0243)	(0.0595)	(0.0589)
Control mean	[-0.7321]	[-0.7321]	[-0.7321]	[-0.7321]
Observations	12,182	11,609	12,182	11,609
US History	-0.0276	-0.0197	-0.0676	-0.0481
	(0.0371)	(0.0378)	(0.0909)	(0.0923)
Control mean	[-0.7902]	[-0.7902]	[-0.7902]	[-0.7902]
Observations	8,559	8,191	8,559	8,191
Retake EOC				
Algebra I	0.0314*	0.0307*	0.0755*	0.0742*
	(0.0174)	(0.0172)	(0.0422)	(0.0418)
Control mean	[0.2919]	[0.2919]	[0.2919]	[0.2919]
Observations	12,666	12,045	12,666	12,045
Biology	-0.0080	-0.0113	-0.0193	-0.0274
	(0.0169)	(0.0174)	(0.0408)	(0.0423)
Control mean	[0.2572]	[0.2572]	[0.2572]	[0.2572]
Observations	12,540	11,934	12,540	11,934
English I	0.0072	0.0083	0.0174	0.0203
	(0.0184)	(0.0179)	(0.0447)	(0.0438)
Control mean	[0.7836]	[0.7836]	[0.7836]	[0.7836]
Observations	12,431	11,817	12,431	11,817
English II	0.0041	0.0017	0.0100	0.0042
	(0.0195)	(0.0193)	(0.0475)	(0.0473)
Control mean	[0.7159]	[0.7159]	[0.7159]	[0.7159]
Observations	11,725	11,176	11,725	11,176
US History	0.0013	0.0065	0.0033	0.0159
	(0.0173)	(0.0175)	(0.0424)	(0.0427)
Control mean	[0.1819]	[0.1819]	[0.1819]	[0.1819]
Observations	8,372	8,019	8,372	8,019

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.25 points), calculated via Calonico et al.'s (2014) bandwidth selection procedure. Robust standard errors, clustered at the eighth-grade school level, in parentheses. Control refers to students who scored below the STAAR cut score. All models include year fixed effects. Cols. 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region. For this analysis, we exclude students who did not take the Algebra I EOC before ninth grade. Observation counts vary due to missing test score data for same exams. ITT = Intent-to-treat. CATE = Complier local average treatment effect. STAAR = State of Texas Assessments of Academic Readiness * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 4. *Effects of Reclassification on Probability of Failing at Least One Class, by Subject*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
English	0.0358*	0.0441**	0.0868*	0.1072**
	(0.0208)	(0.0204)	(0.0507)	(0.0498)
Control mean	[0.5462]	[0.5462]	[0.5462]	[0.5462]
Math	0.0026	0.0029	0.0064	0.0071
	(0.0206)	(0.0204)	(0.0500)	(0.0496)
Control mean	[0.6032]	[0.6032]	[0.6032]	[0.6032]
Science	0.0419**	0.0420**	0.1017**	0.1022**
	(0.0204)	(0.0203)	(0.0497)	(0.0498)
Control mean	[0.5007]	[0.5007]	[0.5007]	[0.5007]
Social Studies	0.0270	0.0336*	0.0654	0.0816*
	(0.0193)	(0.0193)	(0.0471)	(0.0473)
Control mean	[0.4689]	[0.4689]	[0.4689]	[0.4689]
Observations	12,946	12,295	12,946	12,295

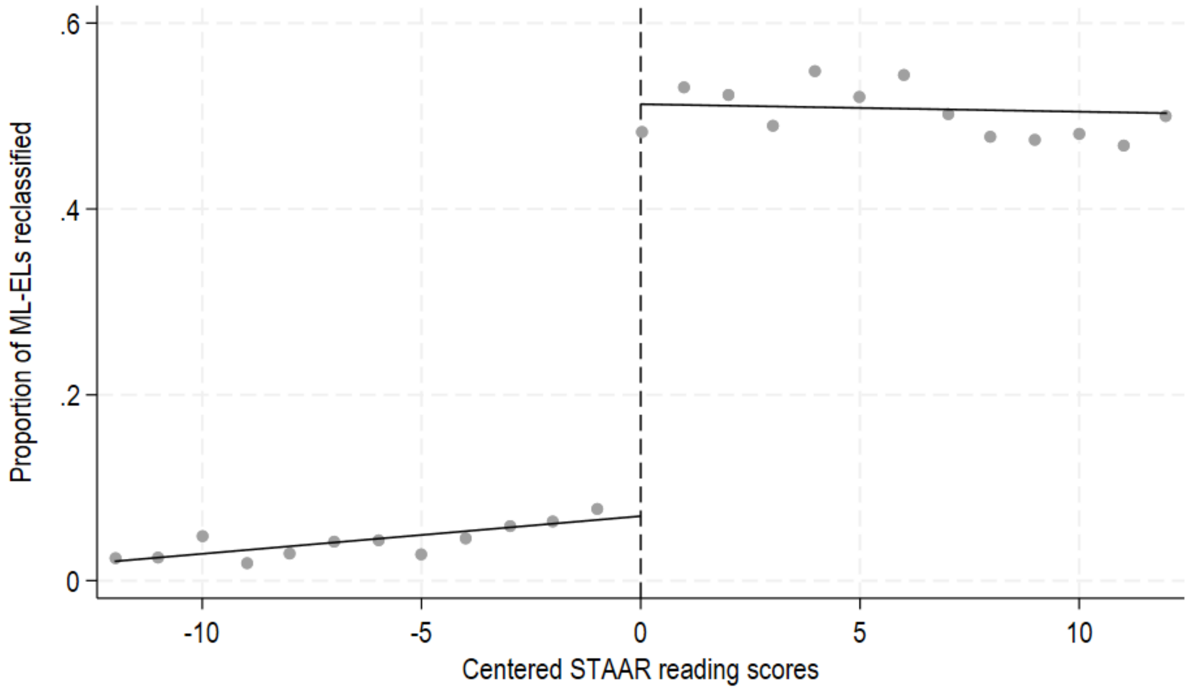
Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.36 points), calculated via Calonico et al.'s (2014) bandwidth selection procedure. Robust standard errors, clustered at the eighth-grade school level, in parentheses. Control refers to students who scored below the STAAR cut score. All models include year fixed effects. Cols. 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region. ITT = Intent-to-treat. CATE = Complier local average treatment effect. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5. *Effects of Reclassification on High School Graduation and Related Outcomes*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Ever Graduated	-0.0199*	-0.0219**	-0.0492*	-0.0543**
	(0.0111)	(0.0108)	(0.0275)	(0.0268)
Control mean	[0.9047]	[0.9047]	[0.9047]	[0.9047]
Observations	11,940	11,325	11,940	11,325
Graduated On-time	-0.0350**	-0.0408***	-0.0867**	-0.1009***
	(0.0151)	(0.0148)	(0.0375)	(0.0370)
Control mean	[0.8332]	[0.8332]	[0.8332]	[0.8332]
Observations	11,940	11,325	11,940	11,325
Graduated College-ready	-0.0257**	-0.0269**	-0.0635**	-0.0665**
	(0.0122)	(0.0124)	(0.0304)	(0.0309)
Control mean	[0.1080]	[0.1080]	[0.1080]	[0.1080]
Observations	11,930	11,322	11,930	11,322
Dropped out	0.0199**	0.0219**	0.0492**	0.0543**
	(0.0116)	(0.0111)	(0.0275)	(0.0268)
Control mean	[0.0953]	[0.0953]	[0.0953]	[0.0953]
Observations	11,940	11,325	11,940	11,325

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.36 points), calculated via Calonico et al.'s (2014) bandwidth selection procedure. Robust standard errors, clustered at the eighth-grade school level, in parentheses. Control refers to students who scored below the STAAR cut score. All models include year fixed effects. Cols. 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region. “On-time” graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including Algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. ITT = Intent-to-treat. CATE = Complier local average treatment effect. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

FIGURE 1. *Probability of Reclassification as a Function of STAAR Reading Score*



Note: The the running variable is raw score on the first attempt of STAAR reading, centered at the cutoff.

Note. Figure shows the probability of being reclassified at the end of eighth grade as a function of students' scores on the second administration of STAAR Reading, conditional on scoring "advanced high" on TELPAS writing, speaking, and listening. STAAR = State of Texas Assessments of Academic Readiness. TELPAS = Texas English Language Proficiency Assessment System. The difference at the threshold is 39 percentage points.

Appendix A: Tables

TABLE A1. *EL Program Placement by Grade, All ELs*

Grade	1	2	3	4	5	6	7	8	9	10	11	12
None	6.3%	6.9%	7.4%	7.7%	8.2%	9.4%	10.0%	10.1%	10.8%	9.3%	8.4%	9.7%
Pull-Out	7.3%	7.3%	8.3%	9.0%	10.5%	42.8%	51.2%	51.8%	38.7%	39.2%	39.6%	38.0%
All Content	16.2%	16.7%	16.4%	16.0%	16.5%	36.5%	37.0%	36.9%	50.3%	51.4%	51.9%	52.3%
Bilingual	70.2%	69.1%	67.8%	67.3%	64.7%	11.2%	1.7%	1.2%	<1%	<1%	<1%	<1%

Note. Descriptive statistics for all EL-classified students enrolled in Texas public schools during the 2012-2013 to 2016-2017 academic years.

TABLE A2. *Descriptive Characteristics of Grade 8 Students in Texas, 2013-2017*

	(1) ML-ELs	(2) Valid scores	(3) Met TELPAS	(4) Took Test 2; no accom.	(5) Enr. in G9	(6) Analytic RD (within 2SD)
Male	55%	55%	49%	52%	52%	51%
Econ. Disadvantage	88%	88%	88%	89%	89%	89%
Home language: Spanish	91%	93%	93%	95%	95%	95%
Race/Ethnicity						
Asian	4%	4%	3%	2%	2%	2%
Black	1%	1%	1%	1%	1%	1%
Hispanic/Latino	92%	93%	94%	95%	96%	95%
Native American	<1%	<1%	<1%	<1%	<1%	<1%
NHPI	<1%	<1%	<1%	<1%	<1%	<1%
Multiracial	<1%	<1%	<1%	<1%	<1%	<1%
White	2%	2%	2%	1%	1%	1%
Years in US Schools						
1-2 years	9%	10%	3%	1%	1%	1%
3-4 years	7%	9%	5%	3%	3%	3%
5+ years	72%	80%	91%	95%	95%	95%
SPED in Grade 8	12%	7%	5%	7%	7%	7%
Age	13.34	13.31	13.25	13.33	13.33	13.32
Repeated Grade 8	1%	<1%	<1%	<1%	<1%	<1%
Enrolled in Grade 9	95%	97%	98%	98%	>99%	>99%
Observations	208,060	169,665	66,712	22,929	22,454	21,044
Met STAAR Standard						
Grade 7 Reading	31%	29%	38%	16%	16%	16%
Grade 7 Math	39%	39%	47%	32%	32%	33%
Grade 7 Writing	26%	25%	36%	19%	19%	20%
Observations	184,274	156,406	64,500	22,434	22,007	20,635
% of All ML-Els	100%	82%	32%	11%	10%	10%

Note. Descriptive statistics for students enrolled in eighth grade in Texas public schools during the 2012-2013 to 2016-2017 academic years. Column 1 shows all English-learner classified eighth-grade students. Column 2 shows ML-ELs who have valid a score on both TELPAS and STAAR reading tests. Column 3 shows students who scored 'advanced high' on the writing, speaking, and listening domains of TELPAS, requirements for reclassification. Column 4 shows students who failed the first STAAR attempt and re-took the second administration of the test without receiving testing accommodations. Column 5 excludes students who did not enroll in a public school in Texas in the following grade. Column 6 restricts the sample to students whose STAAR scores are within 2SD around the cutoff. Observation counts vary due to missing test score data. ML-EL = Multilingual English Learner. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. SPED = Identified for Special Education. Econ. Disadvantage= Economically disadvantaged.

Table A3: *Reclassification Timing Among Students Who Did Not Reclassify in Grade 8*

Reclassified by Grade	Percent
9	12%
10	20%
11	24%
Never	67%
Drop Out	8%

Note. Sample includes students in the sample who did not reclassify in eighth grade and who scored within 5 points of the reclassification threshold. Grade-level cells show the cumulative percentage of students who reclassified by the end of each grade. Never shows the percentage of students who did not reclassify by the end of high school. Drop out shows the percentage of students who dropped out before reclassifying.

TABLE A4. *Estimates of Differences in Observable Characteristics Between Students Who Fall Above and Below the Reclassification Threshold on the STAAR Reading Exam*

	(1) BW4	(2) BW5	(3) BW6	(4) BW7	(5) BW8	(6) BW9	(7) BW10
Student Characteristics							
Male	-0.0167	-0.0204	-0.0182	-0.0181	-0.0144	-0.0137	-0.0121
	(0.0244)	(0.0203)	(0.0179)	(0.0164)	(0.0155)	(0.0146)	(0.0140)
Hispanic	-0.0012	-0.0055	-0.0047	-0.0049	-0.0045	-0.0041	-0.0039
	(0.0113)	(0.0100)	(0.0091)	(0.0085)	(0.0080)	(0.0077)	(0.0074)
Age	-0.0048	-0.0015	0.0038	0.0037	0.0042	0.0037	0.0010
	(0.0251)	(0.0214)	(0.0191)	(0.0174)	(0.0163)	(0.0155)	(0.0150)
Econ. Disad.	0.0073	-0.0053	-0.0078	-0.0074	-0.0076	-0.0060	-0.0044
	(0.0161)	(0.0141)	(0.0126)	(0.0116)	(0.0108)	(0.0103)	(0.0098)
Special Ed.	0.0257**	0.0191**	0.0164*	0.0141*	0.0128*	0.0123*	0.0115*
	(0.0108)	(0.0092)	(0.0085)	(0.0079)	(0.0075)	(0.0072)	(0.0069)
1-2 yrs in U.S. sch	0.0033	0.0019	0.0004	-0.0004	-0.0003	-0.0004	-0.0005
	(0.0054)	(0.0046)	(0.0041)	(0.0038)	(0.0036)	(0.0035)	(0.0033)
3-4 yrs in U.S. sch	-0.0011	-0.0009	-0.0015	-0.0012	-0.0009	-0.0003	0.0002
	(0.0086)	(0.0073)	(0.0064)	(0.0059)	(0.0055)	(0.0052)	(0.0049)
Observations	9154	11453	13440	15210	16712	18039	19047
Met G7 Reading	0.0242	0.0232	0.0209	0.0178	0.0153	0.0137	0.0126
	(0.0184)	(0.0155)	(0.0139)	(0.0128)	(0.0120)	(0.0114)	(0.0109)
	8824	11042	12939	14637	16066	17333	18291
Met G7 Math	-0.0006	-0.0040	-0.0014	-0.0026	-0.0032	-0.0030	-0.0022
	(0.0205)	(0.0175)	(0.0159)	(0.0148)	(0.0140)	(0.0133)	(0.0127)
	8834	11062	12971	14679	16108	17381	18348
Took Algebra I G8	0.0008	0.0019	0.0025	0.0027	0.0028	0.0026	0.0026
	(0.0085)	(0.0073)	(0.0066)	(0.0062)	(0.0058)	(0.0056)	(0.0055)
	9015	11284	13238	14979	16463	17755	18742
Grade 8 school							
Charter school	0.0046	0.0083	0.0061	0.0052	0.0036	0.0032	0.0028
	(0.0106)	(0.0093)	(0.0086)	(0.0082)	(0.0079)	(0.0077)	(0.0075)
Enrollment	3.3734	4.3226	5.3169	5.2253	5.3768	5.7583	6.1870
	(21.8326)	(20.3749)	(19.6444)	(19.2152)	(18.9471)	(18.7027)	(18.4914)
% Econ. Disad.	0.4587	-0.1150	-0.0520	-0.0950	-0.1219	-0.0613	-0.0358
	(1.3271)	(1.2283)	(1.1723)	(1.1348)	(1.1096)	(1.0912)	(1.0753)
% Hispanic	1.3369	0.7802	0.6116	0.2679	0.0986	0.0803	0.0713
	(1.8336)	(1.7247)	(1.6520)	(1.6101)	(1.5845)	(1.5634)	(1.5478)
% White	-0.3046	0.1313	0.1322	0.2742	0.3205	0.2648	0.2538
	(1.1964)	(1.1084)	(1.0560)	(1.0210)	(0.9981)	(0.9815)	(0.9688)
% ML-ELs	0.6353	0.2996	0.4359	0.3677	0.3328	0.3552	0.3986
	(1.1203)	(1.0703)	(1.0364)	(1.0162)	(1.0064)	(0.9987)	(0.9888)
Observations	9154	11453	13440	15210	16712	18039	19047
% Passed Reading G8	0.4403	0.4225	0.3267	0.3537	0.3300	0.3045	0.2899
	(0.7878)	(0.7272)	(0.6872)	(0.6658)	(0.6504)	(0.6409)	(0.6343)
% Passed Math G8	0.8493	0.6862	0.4530	0.2509	0.0695	-0.0355	-0.1068
	(0.8712)	(0.8042)	(0.7648)	(0.7410)	(0.7233)	(0.7109)	(0.7028)
Observations	9141	11433	13412	15171	16668	17990	18991
% Passed Reading G7	-0.3466	-0.0699	-0.1453	-0.1536	-0.1651	-0.1716	-0.1641
	(0.8471)	(0.7858)	(0.7446)	(0.7204)	(0.7054)	(0.6938)	(0.6849)
% Passed Math G7	0.6416	0.6684	0.5621	0.5687	0.4757	0.4026	0.3450
	(1.0675)	(0.9821)	(0.9340)	(0.9046)	(0.8820)	(0.8646)	(0.8523)
	9087	11359	13326	15071	16557	17870	18866

% ML-ELs in district	0.2871	0.0877	0.0806	-0.0616	-0.1543	-0.1772	-0.1710
	(0.9320)	(0.8923)	(0.8686)	(0.8519)	(0.8418)	(0.8356)	(0.8292)
Observations	9154	11453	13440	15210	16712	18039	19047
Grade 9 school							
Enrollment	-46.2158	-29.6889	-18.0863	-16.7746	-16.7959	-16.3402	-14.5510
	(69.2859)	(64.8024)	(62.3891)	(61.0112)	(59.9707)	(59.1414)	(58.3807)
% Econ. Disad.	1.6423	0.9161	0.7096	0.4853	0.3233	0.3188	0.2977
	(1.3414)	(1.2459)	(1.1971)	(1.1603)	(1.1366)	(1.1204)	(1.1092)
% Hispanic	1.7083	1.1848	0.8016	0.4171	0.1979	0.1474	0.0947
	(1.7863)	(1.6798)	(1.6126)	(1.5733)	(1.5510)	(1.5334)	(1.5199)
% White	-0.4835	-0.0981	-0.0416	0.1394	0.2206	0.1731	0.1619
	(1.1872)	(1.0977)	(1.0467)	(1.0117)	(0.9886)	(0.9732)	(0.9622)
% ML-ELs	0.5792	0.4678	0.4564	0.3322	0.2861	0.2882	0.2876
	(0.6721)	(0.6420)	(0.6226)	(0.6109)	(0.6053)	(0.6026)	(0.5991)
Observations	9154	11453	13440	15210	16712	18039	19047
Total AP/IB/DE courses	-0.1363	-0.1434	-0.1273	-0.0400	0.0383	0.0808	0.1059
	(0.6948)	(0.6442)	(0.6206)	(0.6081)	(0.5952)	(0.5879)	(0.5804)
Observations	9142	11436	13421	15188	16687	18012	19016

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and the bandwidth shown on each column title. All columns present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each covariate. All models include year fixed effects. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. The number of observations within the bandwidth are presented the below standard errors. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in ninth grade. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A5. *Running Variable Density Tests by Year, 2013-2017*

Year	T-Stat	P-Value
2013	-1.08	0.28
2014	-0.76	0.44
2015	-0.87	0.38
2016	-0.04	0.97
2017	-0.69	0.49

TABLE A6. *Effects of Reclassification on Student Course Taking, Multiple Bandwidths*

	(1) BW4	(2) BW5	(3) BW6	(4) BW7	(5) BW8	(6) BW9	(7) BW10
AP/IB/DE Courses							
Any course	0.0308**	0.0324**	0.0296**	0.0260**	0.0227**	0.0204**	0.0186*
	(0.0153)	(0.0129)	(0.0116)	(0.0108)	(0.0103)	(0.0099)	(0.0097)
Control mean	[0.1032]	[0.1008]	[0.0973]	[0.0942]	[0.0916]	[0.0896]	[0.0874]
>1 course	0.0199*	0.0202**	0.0195**	0.0176**	0.0157**	0.0146**	0.0136*
	(0.0112)	(0.0095)	(0.0085)	(0.0080)	(0.0076)	(0.0073)	(0.0071)
Control mean	[0.0516]	[0.0506]	[0.0485]	[0.0463]	[0.0452]	[0.0442]	[0.0428]
English	0.0252**	0.0228***	0.0216***	0.0198***	0.0178**	0.0164**	0.0155**
	(0.0101)	(0.0086)	(0.0078)	(0.0073)	(0.0070)	(0.0067)	(0.0066)
Control mean	[0.0419]	[0.0418]	[0.0402]	[0.0383]	[0.0374]	[0.0367]	[0.0357]
Math	0.0063	0.0049	0.0032	0.0024	0.0019	0.0019	0.0018
	(0.0058)	(0.0051)	(0.0046)	(0.0042)	(0.0040)	(0.0038)	(0.0036)
Control mean	[0.0145]	[0.0137]	[0.0126]	[0.0123]	[0.0122]	[0.0119]	[0.0117]
Science	-0.0036	0.0004	0.0013	0.0029	0.0035	0.0044	0.0045
	(0.0077)	(0.0066)	(0.0059)	(0.0054)	(0.0051)	(0.0048)	(0.0046)
Control mean	[0.0206]	[0.0189]	[0.0189]	[0.0186]	[0.0187]	[0.0183]	[0.0176]
Social Studies	0.0317**	0.0271**	0.0246**	0.0218**	0.0197**	0.0181**	0.0166**
	(0.0127)	(0.0108)	(0.0098)	(0.0092)	(0.0088)	(0.0085)	(0.0083)
Control mean	[0.0702]	[0.0690]	[0.0669]	[0.0651]	[0.0630]	[0.0614]	[0.0595]
Remedial Courses							
Math	0.0226	0.0222	0.0193	0.0192	0.0178	0.0164	0.0150
	(0.0186)	(0.0165)	(0.0154)	(0.0144)	(0.0137)	(0.0132)	(0.0128)
Control mean	[0.1977]	[0.2020]	[0.2092]	[0.2115]	[0.2130]	[0.2158]	[0.2182]
Science	0.0055	-0.0009	-0.0082	-0.0088	-0.0126	-0.0136	-0.0137
	(0.0226)	(0.0204)	(0.0191)	(0.0181)	(0.0174)	(0.0168)	(0.0163)
Control mean	[0.3129]	[0.3150]	[0.3210]	[0.3221]	[0.3245]	[0.3273]	[0.3311]
Observations	8,573	10,724	12,564	14,207	15,591	16,813	17,737

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and the bandwidth shown on each column title. All columns present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification—conditional on meeting TELPAS requirements--on each outcome. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. The number of observations within the bandwidth is presented below the Mean row. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. STAAR = State of Texas Assessments of Academic Readiness. TELPAS = Texas English Language Proficiency Assessment System. Remedial Math includes two courses: "Math Models" and "Math Strategies. Remedial Science = Integrated Physics and Chemistry.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A7. *Effects on STAAR End-of-Course (EOC) Achievement, Multiple Bandwidths*

	(1) BW4	(2) BW5	(3) BW6	(4) BW7	(5) BW8	(6) BW9	(7) BW10
Scores (SDs)							
Algebra I	-0.0761**	-0.0576**	-0.0468*	-0.0438*	-0.0388*	-0.0363*	-0.0361*
	(0.0318)	(0.0276)	(0.0250)	(0.0234)	(0.0223)	(0.0212)	(0.0205)
Control mean	[-0.5145]	[-0.5239]	[-0.5377]	[-0.5499]	[-0.5612]	[-0.5734]	[-0.5857]
Observations	8291	10375	12163	13756	15089	16273	17169
Biology	-0.0061	0.0074	0.0152	0.0143	0.0160	0.0161	0.0136
	(0.0318)	(0.0277)	(0.0251)	(0.0233)	(0.0218)	(0.0206)	(0.0198)
Control mean	[-0.7435]	[-0.7523]	[-0.7648]	[-0.7756]	[-0.7862]	[-0.7998]	[-0.8106]
Observations	8225	10291	12069	13650	14973	16143	17026
English I	-0.0278	-0.0192	-0.0114	-0.0087	-0.0083	-0.0082	-0.0083
	(0.0229)	(0.0199)	(0.0181)	(0.0167)	(0.0158)	(0.0150)	(0.0145)
Control mean	[-0.6088]	[-0.6224]	[-0.6354]	[-0.6516]	[-0.6643]	[-0.6757]	[-0.6886]
Observations	8310	10396	12195	13798	15134	16327	17227
English II	-0.0460	-0.0260	-0.0193	-0.0196	-0.0176	-0.0162	-0.0170
	(0.0292)	(0.0254)	(0.0231)	(0.0217)	(0.0206)	(0.0198)	(0.0192)
Control mean	[-0.7166]	[-0.7321]	[-0.7491]	[-0.7623]	[-0.7738]	[-0.7880]	[-0.8017]
Observations	7916	9900	11609	13131	14404	15520	16361
US History	-0.0206	-0.0233	-0.0144	-0.0120	-0.0021	0.0024	0.0038
	(0.0476)	(0.0401)	(0.0355)	(0.0324)	(0.0306)	(0.0292)	(0.0282)
Control mean	[-0.7740]	[-0.7902]	[-0.8096]	[-0.8146]	[-0.8236]	[-0.8353]	[-0.8490]
Observations	5542	6927	8191	9295	10206	11015	11628
Retake EOC							
Algebra I	0.0528**	0.0338*	0.0268	0.0252*	0.0208	0.0188	0.0169
	(0.0212)	(0.0181)	(0.0164)	(0.0152)	(0.0145)	(0.0138)	(0.0132)
Control mean	[0.2853]	[0.2919]	[0.3015]	[0.3064]	[0.3145]	[0.3206]	[0.3286]
Observations	8202	10274	12045	13625	14949	16114	17000
Biology	0.0035	-0.0092	-0.0139	-0.0110	-0.0093	-0.0071	-0.0049
	(0.0215)	(0.0184)	(0.0163)	(0.0150)	(0.0141)	(0.0133)	(0.0127)
Control mean	[0.2508]	[0.2572]	[0.2676]	[0.2770]	[0.2857]	[0.2946]	[0.3020]
Observations	8133	10176	11934	13489	14799	15952	16824
English I	0.0163	0.0105	0.0055	0.0037	0.0032	0.0028	0.0019
	(0.0220)	(0.0189)	(0.0168)	(0.0154)	(0.0144)	(0.0137)	(0.0132)
Control mean	[0.7752]	[0.7836]	[0.7923]	[0.8011]	[0.8083]	[0.8149]	[0.8208]
Observations	8054	10071	11817	13383	14674	15819	16694
English II	0.0079	0.0023	0.0007	-0.0033	-0.0068	-0.0078	-0.0092
	(0.0235)	(0.0203)	(0.0183)	(0.0168)	(0.0158)	(0.0150)	(0.0144)
Control mean	[0.7054]	[0.7159]	[0.7247]	[0.7334]	[0.7402]	[0.7466]	[0.7527]
Observations	7622	9527	11176	12646	13877	14952	15750
US History	0.0003	0.0063	0.0067	0.0070	0.0077	0.0056	0.0041
	(0.0215)	(0.0185)	(0.0166)	(0.0152)	(0.0144)	(0.0137)	(0.0132)
Control mean	[0.1759]	[0.1819]	[0.1910]	[0.1987]	[0.2024]	[0.2071]	[0.2145]
Observations	5438	6784	8019	9098	9988	10764	11353

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and the bandwidth shown on each column title. All columns present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. For this analysis, we further restrict the analytic sample to exclude students who took the Algebra I EOC before ninth grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A8. *Effects of Reclassification on Course Failure, Multiple Bandwidths*

	(1) BW4	(2) BW5	(3) BW6	(4) BW7	(5) BW8	(6) BW9	(7) BW10
Failed English Class	0.0532**	0.0472**	0.0401**	0.0355*	0.0313*	0.0316*	0.0311*
	(0.0239)	(0.0213)	(0.0195)	(0.0181)	(0.0173)	(0.0165)	(0.0160)
Control mean	[0.5396]	[0.5462]	[0.5524]	[0.5577]	[0.5667]	[0.5726]	[0.5794]
Math Class	0.0135	0.0050	0.0003	-0.0033	-0.0032	-0.0021	-0.0011
	(0.0241)	(0.0212)	(0.0195)	(0.0184)	(0.0176)	(0.0168)	(0.0162)
Control mean	[0.5997]	[0.6032]	[0.6059]	[0.6124]	[0.6171]	[0.6221]	[0.6281]
Science Class	0.0451*	0.0444**	0.0390**	0.0389**	0.0394**	0.0413**	0.0434***
	(0.0249)	(0.0213)	(0.0193)	(0.0181)	(0.0173)	(0.0166)	(0.0161)
Control mean	[0.4972]	[0.5007]	[0.5087]	[0.5180]	[0.5254]	[0.5335]	[0.5407]
Social Studies	0.0359	0.0332	0.0342*	0.0308*	0.0270	0.0275*	0.0284*
	(0.0238)	(0.0202)	(0.0183)	(0.0172)	(0.0165)	(0.0158)	(0.0154)
Control mean	[0.4608]	[0.4689]	[0.4731]	[0.4788]	[0.4873]	[0.4944]	[0.5004]
Observations	8,382	10,483	12,295	13,911	15,256	16,464	17,376

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and the bandwidth shown on each column title. All columns present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A9. *Effects of Reclassification on High School Graduation, Multiple Bandwidths*

	(1) BW4	(2) BW5	(3) BW6	(4) BW7	(5) BW8	(6) BW9	(7) BW10
Ever Graduated	-0.0300**	-0.0247**	-0.0198*	-0.0179*	-0.0162*	-0.0160*	-0.0160**
	(0.0137)	(0.0116)	(0.0103)	(0.0094)	(0.0089)	(0.0085)	(0.0082)
Control mean	[0.9072]	[0.9047]	[0.9010]	[0.8978]	[0.8935]	[0.8898]	[0.8866]
Observations	7,731	9,659	11,325	12,793	14,024	15,105	15,926
Graduated On-time	-0.0557***	-0.0434***	-0.0387***	-0.0370***	-0.0356***	-0.0340***	-0.0317***
	(0.0183)	(0.0158)	(0.0143)	(0.0132)	(0.0124)	(0.0118)	(0.0113)
Control mean	[0.8398]	[0.8332]	[0.8257]	[0.8192]	[0.8134]	[0.8093]	[0.8042]
Observations	7,731	9,659	11,325	12,793	14,024	15,105	15,926
Graduated College-ready	-0.0276*	-0.0251**	-0.0263**	-0.0276***	-0.0284***	-0.0279***	-0.0273***
	(0.0145)	(0.0122)	(0.0108)	(0.0099)	(0.0094)	(0.0089)	(0.0085)
Control mean	[0.0974]	[0.0954]	[0.0941]	[0.0926]	[0.0917]	[0.0915]	[0.0906]
Observations	8,573	10,724	12,564	14,207	15,591	16,813	17,737
Dropped out	0.0365***	0.0296**	0.0231**	0.0206**	0.0180*	0.0168*	0.0160*
	(0.0140)	(0.0119)	(0.0106)	(0.0098)	(0.0092)	(0.0088)	(0.0084)
Control mean	[0.1014]	[0.1038]	[0.1077]	[0.1107]	[0.1148]	[0.1182]	[0.1219]
Observations	7,781	9,722	11,401	12,880	14,117	15,204	16,030

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and the bandwidth shown on each column title. All Columns present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. On-time' graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including Algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A10. *Effects of Reclassification on Student Course Taking, Falsification Tests*

	(1) -4/+Year FE	(2) -4/+Pretreatment Covariates	(3) +4/+Year FE	(4) +4/+Pretreatment Covariates
AP/IB/DE courses				
Any course	0.0164	0.0118	0.0059	0.0028
	(0.0170)	(0.0163)	(0.0121)	(0.0124)
Control mean	[0.1008]	[0.1008]	[0.1008]	[0.1008]
>1 course	0.0144	0.0135	-0.0001	-0.0031
	(0.0124)	(0.0123)	(0.0090)	(0.0093)
Control mean	[0.0506]	[0.0506]	[0.0506]	[0.0506]
English	0.0079	0.0055	-0.0039	-0.0055
	(0.0120)	(0.0119)	(0.0086)	(0.0089)
Control mean	[0.0418]	[0.0418]	[0.0418]	[0.0418]
Math	0.0090	0.0089	0.0055	0.0053
	(0.0065)	(0.0063)	(0.0041)	(0.0043)
Control mean	[0.0137]	[0.0137]	[0.0137]	[0.0137]
Science	0.0059	0.0046	0.0078	0.0058
	(0.0068)	(0.0069)	(0.0051)	(0.0054)
Control mean	[0.0189]	[0.0189]	[0.0189]	[0.0189]
Social Studies	0.0098	0.0073	-0.0034	-0.0054
	(0.0147)	(0.0144)	(0.0103)	(0.0107)
Control mean	[0.0690]	[0.0690]	[0.0690]	[0.0690]
Remedial courses				
Math	-0.0125	-0.0139	-0.0095	-0.0074
	(0.0166)	(0.0168)	(0.0175)	(0.0179)
Control mean	[0.2020]	[0.2020]	[0.2020]	[0.2020]
Science	0.0108	0.0072	0.0088	0.0150
	(0.0225)	(0.0213)	(0.0242)	(0.0223)
Control mean	[0.3150]	[0.3150]	[0.3150]	[0.3150]
Observations	9,690	9,131	13,429	12,505

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.355 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. In this analysis we use two alternative thresholds for students' reading scores All Columns present intent-to-treat (ITT) estimates, showing the impact of passing the alternative STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. In Columns 1 and 2 the alternative threshold is 4 points below the STAAR passing standard and in Columns 2 and 4 the alternative threshold is 4 points above the STAAR passing standard. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the optimal bandwidth. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. FE = Fixed effects. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: “Math Models” and “Math Strategies. Remedial Science = Integrated Physics and Chemistry. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A11. *Effects of Reclassification on STAAR End-of-Course Achievement, Falsification Tests*

	(1) -4/Year FE	(2) -4/+Pretreatment Covariates	(3) +4/Year FE	(4) +4/+Pretreatment Covariates
EOC Scores (SDs)				
Algebra I	0.0393	0.0325	-0.0089	-0.0161
	(0.0284)	(0.0274)	(0.0279)	(0.0277)
Control mean	[-0.5239]	[-0.5239]	[-0.5239]	[-0.5239]
Observations	9161	8751	12907	12198
Biology	0.0294	0.0144	-0.0157	-0.0238
	(0.0262)	(0.0256)	(0.0260)	(0.0264)
Control mean	[-0.7523]	[-0.7523]	[-0.7523]	[-0.7523]
Observations	9100	8696	12776	12088
English I	0.0262	0.0118	-0.0122	-0.0240
	(0.0204)	(0.0194)	(0.0208)	(0.0208)
Control mean	[-0.6224]	[-0.6224]	[-0.6224]	[-0.6224]
Observations	9182	8770	12964	12250
English II	-0.0225	-0.0341	0.0250	0.0081
	(0.0304)	(0.0301)	(0.0284)	(0.0278)
Control mean	[-0.7321]	[-0.7321]	[-0.7321]	[-0.7321]
Observations	8797	8419	12207	11566
US History	-0.0050	-0.0304	0.0291	0.0074
	(0.0443)	(0.0434)	(0.0418)	(0.0416)
Control mean	[-0.7902]	[-0.7902]	[-0.7902]	[-0.7902]
Observations	6059	5816	8760	8343
Retake EOC				
Algebra I	-0.0173	-0.0175	0.0012	0.0032
	(0.0190)	(0.0188)	(0.0198)	(0.0198)
Control mean	[0.2919]	[0.2919]	[0.2919]	[0.2919]
Observations	9069	8664	12770	12073
Biology	-0.0012	-0.0003	0.0108	0.0175
	(0.0177)	(0.0176)	(0.0192)	(0.0194)
Control mean	[0.2572]	[0.2572]	[0.2572]	[0.2572]
Observations	9005	8606	12604	11928
English I	-0.0340	-0.0116	0.0099	0.0086
	(0.0224)	(0.0222)	(0.0165)	(0.0166)
Control mean	[0.7836]	[0.7836]	[0.7836]	[0.7836]
Observations	8936	8534	12538	11849
English II	-0.0140	0.0000	-0.0092	-0.0129
	(0.0240)	(0.0242)	(0.0193)	(0.0195)
Control mean	[0.7159]	[0.7159]	[0.7159]	[0.7159]
Observations	8507	8143	11709	11103
US History	0.0151	0.0123	-0.0054	-0.0020
	(0.0167)	(0.0167)	(0.0211)	(0.0211)
Control mean	[0.1819]	[0.1819]	[0.1819]	[0.1819]
Observations	5952	5717	8517	8117

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth (BW = 5.25 points), selected via Calonico et al.'s (2014) bandwidth selection procedure. In this analysis we use two alternative thresholds for students' reading scores. All columns present intent-to-treat (ITT) estimates. In Columns 1 and 2, the alternative threshold is 4 points below the STAAR cutoff and in Columns 2 and 4 the alternative threshold is 4 points above the STAAR cutoff. All models include year fixed effects. Columns 2 and 4 add covariates for student and eighth-grade school characteristics, and district region within Texas. Robust standard errors, in parentheses, are clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold. For this analysis, we further restrict the analytic sample to exclude students who took the Algebra I EOC before ninth grade. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A12. *Effects of Reclassification on Course Failure, Falsification Tests*

	(1) -4/Year FE	(2) -4/+Pretreatment Covariates	(3) +4/Year FE	(4) +4/+Pretreatment Covariates
Failed English Class	-0.0384*	-0.0349	0.0065	0.0163
	(0.0230)	(0.0235)	(0.0216)	(0.0216)
Control mean	[0.5462]	[0.5462]	[0.5462]	[0.5462]
Math Class	-0.0154	-0.0131	0.0127	0.0164
	(0.0238)	(0.0241)	(0.0206)	(0.0207)
Control mean	[0.6032]	[0.6032]	[0.6032]	[0.6032]
Science Class	-0.0276	-0.0258	0.0256	0.0314
	(0.0243)	(0.0245)	(0.0218)	(0.0222)
Control mean	[0.5007]	[0.5007]	[0.5007]	[0.5007]
Social Studies Class	-0.0569**	-0.0556**	-0.0030	0.0016
	(0.0226)	(0.0227)	(0.0225)	(0.0226)
Control mean	[0.4689]	[0.4689]	[0.4689]	[0.4689]
Observations	9,265	8,841	13,094	12,361

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.25 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. In this analysis we use two alternative thresholds for students' reading scores All Columns present intent-to-treat (ITT) estimates, showing the impact of passing the alternative STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. In Columns 1 and 2 the alternative threshold is 4 points below the STAAR passing standard and in Columns 2 and 4 the alternative threshold is 4 points above the STAAR passing standard. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the optimal bandwidth. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. For this analysis, we further restrict the analytic sample to exclude students took the Algebra I EOC before ninth grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A13. *Effects of Reclassification on High School Graduation, Falsification Tests*

	(1) -4/Year FE	(2) -4/+Pretreatment Covariates	(3) +4/Year FE	(4) +4/+Pretreatment Covariates
Ever Graduated	-0.0032	-0.0007	-0.0100	-0.0170
	(0.0124)	(0.0120)	(0.0134)	(0.0129)
Control mean	[0.9047]	[0.9047]	[0.9047]	[0.9047]
Observations	8750	8324	11792	11120
Graduated On-time	-0.0028	-0.0006	0.0140	0.0065
	(0.0166)	(0.0162)	(0.0173)	(0.0167)
Control mean	[0.8332]	[0.8332]	[0.8332]	[0.8332]
Observations	8750	8324	11792	11120
Graduated College-ready	0.0147	0.0097	0.0140	0.0168
	(0.0129)	(0.0132)	(0.0119)	(0.0123)
Control mean	[0.0954]	[0.0954]	[0.0954]	[0.0954]
Observations	9690	9131	13429	12505
Dropped out	0.0041	-0.0013	0.0109	0.0171
	(0.0127)	(0.0122)	(0.0137)	(0.0132)
Control mean	[0.1038]	[0.1038]	[0.1038]	[0.1038]
Observations	8802	8362	11897	11205

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.355 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. In this analysis we use two alternative thresholds for students' reading scores All Columns present intent-to-treat (ITT) estimates, showing the impact of passing the alternative STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. In Columns 1 and 2 the alternative threshold is 4 points below the STAAR passing standard and in Columns 2 and 4 the alternative threshold is 4 points above the STAAR passing standard. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the optimal bandwidth. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. On-time' graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A14. *Effects of Reclassification on High School Coursetaking, Excluding COVID Cohorts 2016-2017*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
AP/IB/DE courses				
Any course	0.0198	0.0216	0.0520	0.0568
	(0.0168)	(0.0164)	(0.0443)	(0.0430)
Control mean	[0.1059]	[0.1059]	[0.1059]	[0.1059]
>1 course	0.0154	0.0153	0.0406	0.0401
	(0.0128)	(0.0128)	(0.0339)	(0.0337)
Control mean	[0.0536]	[0.0536]	[0.0536]	[0.0536]
English	0.0214*	0.0207*	0.0563*	0.0544*
	(0.0115)	(0.0116)	(0.0307)	(0.0307)
Control mean	[0.0460]	[0.0460]	[0.0460]	[0.0460]
Math	-0.0002	0.0008	-0.0007	0.0022
	(0.0086)	(0.0083)	(0.0225)	(0.0219)
Control mean	[0.0211]	[0.0211]	[0.0211]	[0.0211]
Science	-0.0000	-0.0012	-0.0000	-0.0031
	(0.0091)	(0.0093)	(0.0239)	(0.0244)
Control mean	[0.0226]	[0.0226]	[0.0226]	[0.0226]
Social Studies	0.0201	0.0237*	0.0529	0.0622*
	(0.0133)	(0.0133)	(0.0353)	(0.0352)
Control mean	[0.0664]	[0.0664]	[0.0664]	[0.0664]
Remedial courses				
Math	0.0123	0.0172	0.0324	0.0451
	(0.0264)	(0.0261)	(0.0696)	(0.0688)
Control mean	[0.3345]	[0.3345]	[0.3345]	[0.3345]
Science	-0.0332	-0.0375	-0.0873	-0.0985
	(0.0260)	(0.0246)	(0.0687)	(0.0650)
Control mean	[0.3347]	[0.3347]	[0.3347]	[0.3347]
Observations	7,243	6,837	7,243	6,837

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 6.038 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2015; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. This analysis also excludes the pandemic-affected cohorts (2016, 2017). AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: "Math Models" and "Math Strategies. Remedial Science = Integrated Physics and Chemistry.

$^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A15. *Effects of Reclassification on High School Achievement, Excluding COVID Cohorts 2016-2017*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
STAAR EOC Achievement (SDs)				
Algebra I	-0.0580*	-0.0507	-0.1531*	-0.1335
	(0.0346)	(0.0340)	(0.0923)	(0.0905)
Control mean	[-0.5602]	[-0.5602]	[-0.5602]	[-0.5602]
Observations	6085	5840	6085	5840
Biology	-0.0330	-0.0146	-0.0875	-0.0385
	(0.0329)	(0.0325)	(0.0876)	(0.0862)
Control mean	[-0.7118]	[-0.7118]	[-0.7118]	[-0.7118]
Observations	6041	5802	6041	5802
English I	-0.0186	-0.0234	-0.0492	-0.0616
	(0.0240)	(0.0235)	(0.0635)	(0.0621)
Control mean	[-0.6271]	[-0.6271]	[-0.6271]	[-0.6271]
Observations	6108	5860	6108	5860
English II	-0.0536*	-0.0406	-0.1424*	-0.1072
	(0.0322)	(0.0312)	(0.0862)	(0.0828)
Control mean	[-0.7165]	[-0.7165]	[-0.7165]	[-0.7165]
Observations	5810	5592	5810	5592
US History	-0.0460	-0.0324	-0.1249	-0.0865
	(0.0468)	(0.0467)	(0.1279)	(0.1248)
Control mean	[-0.7778]	[-0.7778]	[-0.7778]	[-0.7778]
Observations	5453	5253	5453	5253
Retake EOC				
Algebra I	0.0450*	0.0409	0.1185*	0.1074
	(0.0259)	(0.0255)	(0.0687)	(0.0676)
Control mean	[0.3701]	[0.3701]	[0.3701]	[0.3701]
Observations	6025	5782	6025	5782
Biology	0.0159	0.0132	0.0421	0.0349
	(0.0232)	(0.0237)	(0.0616)	(0.0628)
Control mean	[0.2463]	[0.2463]	[0.2463]	[0.2463]
Observations	5973	5738	5973	5738
English I	0.0262	0.0298	0.0688	0.0782
	(0.0239)	(0.0233)	(0.0628)	(0.0614)
Control mean	[0.7948]	[0.7948]	[0.7948]	[0.7948]
Observations	6023	5779	6023	5779
English II	0.0101	0.0088	0.0272	0.0237
	(0.0253)	(0.0247)	(0.0678)	(0.0660)
Control mean	[0.7583]	[0.7583]	[0.7583]	[0.7583]
Observations	5648	5437	5648	5437
US History	0.0115	0.0146	0.0314	0.0388
	(0.0205)	(0.0206)	(0.0560)	(0.0552)
Control mean	[0.1835]	[0.1835]	[0.1835]	[0.1835]
Observations	5360	5169	5360	5169

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth (BW = 5.67 points) (Calonico et al. 2014). All models include year fixed effects. Cols. 2 and 4 add covariates for student and eighth-grade school characteristics, and district region. Robust standard errors, in parentheses, clustered at the eighth-grade school level. Control means, in brackets, calculated using students who scored below the STAAR threshold. Number of observations within the bandwidth is presented below the Mean row. This analysis excludes pandemic-affected cohorts (2016, 2017). ITT = Intent-to-treat. CATE = Complier average treatment effect. STAAR = State of Texas Assessments of Academic Readiness. EOC = End-of-course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A16. *Effects of Reclassification on Course Completion, Excluding COVID Cohorts 2016-2017*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) Pretreatment Covariates
Failed English Class	0.0451	0.0500*	0.1195	0.1320*
	(0.0287)	(0.0278)	(0.0763)	(0.0737)
Control mean	[0.5702]	[0.5702]	[0.5702]	[0.5702]
Math Class	0.0025	0.0058	0.0065	0.0152
	(0.0277)	(0.0272)	(0.0735)	(0.0718)
Control mean	[0.6302]	[0.6302]	[0.6302]	[0.6302]
Science Class	0.0545*	0.0621**	0.1444*	0.1639**
	(0.0284)	(0.0277)	(0.0757)	(0.0737)
Control mean	[0.5209]	[0.5209]	[0.5209]	[0.5209]
Social Studies Class	0.0795***	0.0897***	0.2106***	0.2367***
	(0.0269)	(0.0263)	(0.0730)	(0.0718)
Control mean	[0.4841]	[0.4841]	[0.4841]	[0.4841]
Observations	6,166	5,906	6,166	5,906

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.67 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2015; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. For this analysis, we further restrict the analytic sample to exclude students who did not take the Algebra I EOC before ninth grade. This analysis also excludes the pandemic-affected cohorts (2016, 2017). TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A17. *Effects of Reclassification on High School Graduation, Excluding COVID Cohorts 2016-2017*

	ITT		CATE	
	(1) Year FE	(2) +Pretreatment Covariates	(3) Year FE	(4) +Pretreatment Covariates
Ever Graduated	-0.0056	-0.0092	-0.0153	-0.0249
	(0.0159)	(0.0158)	(0.0434)	(0.0427)
Control mean	[0.8929]	[0.8929]	[0.8929]	[0.8929]
Observations	6347	6075	6347	6075
Graduated On-time	-0.0397*	-0.0446**	-0.1081*	-0.1206**
	(0.0205)	(0.0202)	(0.0559)	(0.0550)
Control mean	[0.8176]	[0.8176]	[0.8176]	[0.8176]
Observations	6347	6075	6347	6075
Graduated College-ready	-0.0444**	-0.0466**	-0.1170**	-0.1225**
	(0.0185)	(0.0189)	(0.0494)	(0.0504)
Control mean	[0.1364]	[0.1364]	[0.1364]	[0.1364]
Observations	7243	6837	7243	6837
Dropped out	0.0051	0.0114	0.0136	0.0306
	(0.0165)	(0.0161)	(0.0445)	(0.0432)
Control mean	[0.1171]	[0.1171]	[0.1171]	[0.1171]
Observations	6402	6123	6402	6123

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 6.04 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2015; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. This analysis also excludes the pandemic-affected cohorts (2016-2017). On-time' graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including Algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A18. *Effects of Reclassification on the Probability of Having Nonmissing Data on Student Coursetaking Outcomes*

	ITT		CATE	
	(1) +Year Fe	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
AP/IB/DE courses				
Any course	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
>1 course	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
English	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Math	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Science	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Social Studies	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Remedial courses				
Math	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Integrated Physics & Chemistry	0.0033	0.0046	0.0079	0.0043
	(0.0028)	(0.0028)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Observations	13440	12754	13440	12577

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.355 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on having nonmissing information on each variable. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on the probability of having nonmissing information on each variable for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A19. *Effects of Reclassification on the Probability of Having Nonmissing Data on Student Achievement Outcomes*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
STAAR EOC Achievement (SDs)				
Algebra I	-0.0078	-0.0062	-0.0189	-0.0152
	(0.0048)	(0.0047)	(0.0116)	(0.0115)
Control mean	[0.9835]	[0.9835]	[0.9835]	[0.9835]
Biology	-0.0051	-0.0036	-0.0124	-0.0087
	(0.0058)	(0.0055)	(0.0140)	(0.0135)
Control mean	[0.9733]	[0.9733]	[0.9733]	[0.9733]
English I	-0.0037	-0.0020	-0.0091	-0.0049
	(0.0045)	(0.0044)	(0.0109)	(0.0107)
Control mean	[0.9868]	[0.9868]	[0.9868]	[0.9868]
English II	-0.0065	-0.0051	-0.0158	-0.0125
	(0.0095)	(0.0094)	(0.0232)	(0.0229)
Control mean	[0.9282]	[0.9282]	[0.9282]	[0.9282]
US History	-0.0080	-0.0072	-0.0195	-0.0175
	(0.0144)	(0.0141)	(0.0352)	(0.0344)
Control mean	[0.6656]	[0.6656]	[0.6656]	[0.6656]
Retake EOC				
Algebra I	-0.0095	-0.0068	-0.0231	-0.0166
	(0.0064)	(0.0063)	(0.0157)	(0.0155)
Control mean	[0.9741]	[0.9741]	[0.9741]	[0.9741]
Biology	-0.0074	-0.0044	-0.0179	-0.0107
	(0.0074)	(0.0071)	(0.0180)	(0.0173)
Control mean	[0.9616]	[0.9616]	[0.9616]	[0.9616]
English I	-0.0092	-0.0050	-0.0224	-0.0123
	(0.0100)	(0.0094)	(0.0242)	(0.0230)
Control mean	[0.9550]	[0.9550]	[0.9550]	[0.9550]
English II	-0.0184	-0.0145	-0.0447	-0.0353
	(0.0134)	(0.0127)	(0.0326)	(0.0309)
Control mean	[0.8912]	[0.8912]	[0.8912]	[0.8912]
US History	-0.0111	-0.0096	-0.0270	-0.0234
	(0.0148)	(0.0145)	(0.0360)	(0.0355)
Control mean	[0.6494]	[0.6494]	[0.6494]	[0.6494]
Observations	12,995	12,339	12,995	12,339

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.25 points), selected via Calonico et al.'s (2014) bandwidth selection procedure. Cols. 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on having nonmissing information on each variable. Cols. 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on the probability of having nonmissing information on each variable for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Cols. 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. For this analysis, we restrict the analytic sample to exclude students who did not take the Algebra I EOC before ninth grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A20. *Effects of Reclassification on the Probability of Having Nonmissing Data on Course Failure Outcomes*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Failed English Class	0.0033	0.0047	0.0081	0.0114
	(0.0029)	(0.0030)	(0.0072)	(0.0072)
Control mean	[0.9955]	[0.9955]	[0.9955]	[0.9955]
Math Class	0.0033	0.0047	0.0081	0.0114
	(0.0029)	(0.0030)	(0.0072)	(0.0072)
Control mean	[0.9955]	[0.9955]	[0.9955]	[0.9955]
Science Class	0.0033	0.0047	0.0081	0.0114
	(0.0029)	(0.0030)	(0.0072)	(0.0072)
Control mean	[0.9955]	[0.9955]	[0.9955]	[0.9955]
Social Studies Class	0.0033	0.0047	0.0081	0.0114
	(0.0029)	(0.0030)	(0.0072)	(0.0072)
Control mean	[0.9955]	[0.9955]	[0.9955]	[0.9955]
Observations	12,995	12,339	12,995	12,339

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.25 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on having nonmissing information on each variable. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on the probability of having nonmissing information on each variable for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. For this analysis, we further restrict the analytic sample to exclude students who did not take the Algebra I EOC before ninth grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A21. *Effects of Reclassification on the Probability of Having Nonmissing Data on High School Graduation Outcomes*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Ever Graduated	-0.0036	0.0008	-0.0088	0.0019
	(0.0127)	(0.0121)	(0.0309)	(0.0295)
Control mean	[0.8759]	[0.8759]	[0.8759]	[0.8759]
Graduated On-time	-0.0036	0.0008	-0.0088	0.0019
	(0.0127)	(0.0121)	(0.0309)	(0.0295)
Control mean	[0.8759]	[0.8759]	[0.8759]	[0.8759]
Graduated College-ready	0.0033	0.0018	0.0079	0.0043
	(0.0028)	(0.0016)	(0.0069)	(0.0038)
Control mean	[0.9956]	[0.9956]	[0.9956]	[0.9956]
Dropped out	-0.0003	0.0046	-0.0007	0.0113
	(0.0124)	(0.0118)	(0.0302)	(0.0286)
Control mean	[0.8843]	[0.8843]	[0.8843]	[0.8843]
Observations	13,440	12,577	13,440	12,577

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.355 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on having nonmissing information on each variable. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on the probability of having nonmissing information on each variable for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. On-time' graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A22. *Effects of Reclassification on Student Course Taking (Stable Sample)*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
AP/IB/DE courses				
Any course	0.0566***	0.0502***	0.1416***	0.1249***
	(0.0167)	(0.0163)	(0.0427)	(0.0410)
Control mean	[0.1217]	[0.1217]	[0.1217]	[0.1217]
>1 course	0.0354***	0.0315**	0.0887***	0.0785**
	(0.0126)	(0.0124)	(0.0320)	(0.0313)
Control mean	[0.0624]	[0.0624]	[0.0624]	[0.0624]
English	0.0361***	0.0314***	0.0904***	0.0781***
	(0.0116)	(0.0115)	(0.0294)	(0.0288)
Control mean	[0.0542]	[0.0542]	[0.0542]	[0.0542]
Math	0.0076	0.0085	0.0189	0.0212
	(0.0078)	(0.0076)	(0.0195)	(0.0191)
Control mean	[0.0213]	[0.0213]	[0.0213]	[0.0213]
Science	0.0051	0.0021	0.0128	0.0051
	(0.0090)	(0.0092)	(0.0227)	(0.0228)
Control mean	[0.0267]	[0.0267]	[0.0267]	[0.0267]
Social Studies	0.0503***	0.0467***	0.1261***	0.1163***
	(0.0133)	(0.0133)	(0.0345)	(0.0339)
Control mean	[0.0755]	[0.0755]	[0.0755]	[0.0755]
Remedial courses				
Math	0.0221	0.0268	0.0554	0.0668
	(0.0221)	(0.0218)	(0.0557)	(0.0548)
Control mean	[0.2551]	[0.2551]	[0.2551]	[0.2551]
Science	-0.0173	-0.0255	-0.0433	-0.0634
	(0.0257)	(0.0242)	(0.0644)	(0.0604)
Control mean	[0.3393]	[0.3393]	[0.3393]	[0.3393]
Observations	7,651	7,336	7,651	7,336

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.574 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. For this estimation, we restrict the analytic sample to students with nonmissing information on all outcomes (stable sample). AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: “Math Models” and “Math Strategies. Remedial Science = Integrated Physics and Chemistry. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A23. *Effects of Reclassification on Student Achievement Outcomes (Stable Sample)*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
STAAR EOC Achievement (SDs)				
Algebra I	0.0030	0.0035	0.0074	0.0087
	(0.0301)	(0.0296)	(0.0751)	(0.0734)
Control mean	[-0.5055]	[-0.5055]	[-0.5055]	[-0.5055]
Biology	0.0343	0.0404	0.0857	0.1001
	(0.0269)	(0.0269)	(0.0673)	(0.0665)
Control mean	[-0.6773]	[-0.6773]	[-0.6773]	[-0.6773]
English I	-0.0052	-0.0115	-0.0131	-0.0286
	(0.0202)	(0.0198)	(0.0505)	(0.0492)
Control mean	[-0.6192]	[-0.6192]	[-0.6192]	[-0.6192]
English II	0.0129	0.0106	0.0321	0.0263
	(0.0265)	(0.0252)	(0.0662)	(0.0624)
Control mean	[-0.6845]	[-0.6845]	[-0.6845]	[-0.6845]
US History	-0.0139	-0.0085	-0.0347	-0.0210
	(0.0357)	(0.0359)	(0.0894)	(0.0890)
Control mean	[-0.7279]	[-0.7279]	[-0.7279]	[-0.7279]
Retake EOC				
Algebra I	0.0153	0.0127	0.0381	0.0315
	(0.0224)	(0.0224)	(0.0562)	(0.0556)
Control mean	[0.3025]	[0.3025]	[0.3025]	[0.3025]
Biology	-0.0295	-0.0307	-0.0738	-0.0762
	(0.0210)	(0.0212)	(0.0527)	(0.0527)
Control mean	[0.2438]	[0.2438]	[0.2438]	[0.2438]
English I	0.0149	0.0216	0.0373	0.0537
	(0.0217)	(0.0211)	(0.0543)	(0.0526)
Control mean	[0.8094]	[0.8094]	[0.8094]	[0.8094]
English II	0.0022	0.0058	0.0054	0.0144
	(0.0228)	(0.0223)	(0.0569)	(0.0552)
Control mean	[0.7607]	[0.7607]	[0.7607]	[0.7607]
US History	-0.0011	0.0025	-0.0026	0.0063
	(0.0174)	(0.0176)	(0.0434)	(0.0437)
Control mean	[0.1806]	[0.1806]	[0.1806]	[0.1806]
Observations	7,397	7,103	7,397	7,103

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.588 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. For this estimation, we restrict the analytic sample to students with nonmissing information on all outcomes (stable sample). TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. EOC = End-of-course.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A24. *Effects of Reclassification on High School Course Failure (Stable Sample)*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Failed English Class	0.0321	0.0369	0.0801	0.0916
	(0.0261)	(0.0254)	(0.0656)	(0.0633)
Control mean	[0.5361]	[0.5361]	[0.5361]	[0.5361]
Math Class	-0.0194	-0.0161	-0.0483	-0.0400
	(0.0266)	(0.0262)	(0.0663)	(0.0648)
Control mean	[0.6029]	[0.6029]	[0.6029]	[0.6029]
Science Class	0.0505*	0.0547**	0.1260*	0.1356**
	(0.0260)	(0.0256)	(0.0657)	(0.0644)
Control mean	[0.4745]	[0.4745]	[0.4745]	[0.4745]
Social Studies Class	0.0415*	0.0510**	0.1036*	0.1265**
	(0.0247)	(0.0241)	(0.0624)	(0.0606)
Control mean	[0.4489]	[0.4489]	[0.4489]	[0.4489]
Observations	7,397	7,103	7,397	7,103

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.588 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. For this analysis, we further restrict the analytic sample to exclude students who did not take the Algebra I EOC before ninth grade. For this estimation, we further restrict the analytic sample to students with nonmissing information on all outcomes (stable sample). TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A25. *Effects of Reclassification on High School Graduation (Stable Sample)*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Ever Graduated	-0.0108	-0.0161	-0.0271	-0.0402
	(0.0103)	(0.0103)	(0.0258)	(0.0256)
Control mean	[0.9466]	[0.9466]	[0.9466]	[0.9466]
Graduated On-time	-0.0384**	-0.0443***	-0.0962**	-0.1103***
	(0.0156)	(0.0154)	(0.0391)	(0.0387)
Control mean	[0.8742]	[0.8742]	[0.8742]	[0.8742]
Graduated College-ready	-0.0376**	-0.0378**	-0.0941**	-0.0940**
	(0.0179)	(0.0180)	(0.0453)	(0.0451)
Control mean	[0.1550]	[0.1550]	[0.1550]	[0.1550]
Dropped out	0.0108	0.0161	0.0271	0.0402
	(0.0103)	(0.0103)	(0.0258)	(0.0256)
Control mean	[0.0534]	[0.0534]	[0.0534]	[0.0534]
Observations	7,651	7,336	7,651	7,336

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.574 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A26. *Effects of Reclassification on the Probability of Enrolling in Ninth Grade*

	ITT		CATE	
	(1) +Year FE	(2) +Pretreatment Covariates	(3) +Year FE	(4) +Pretreatment Covariates
Enrolled in Grade 9	0.0029	-0.0008	0.0072	-0.0019
	(0.0052)	(0.0034)	(0.0129)	(0.0083)
Control mean	[0.9800]	[0.9800]	[0.9800]	[0.9800]
Observations	15,526	14,333	15,526	14,333

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 6.232 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR and 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A27. *Effects of Reclassification on Student Course Taking by G9 EL Program*

	ITT		CATE	
	(1) ESL ELA	(2) ESL All-Content	(3) ESL ELA	(4) ESL All-Content
AP/IB/DE courses				
Any course	0.0213	0.0377*	0.0456	0.1159*
	(0.0161)	(0.0210)	(0.0348)	(0.0653)
Control mean	[0.0975]	[0.1045]	[0.0975]	[0.1045]
>1 course	0.0205*	0.0200	0.0438*	0.0614
	(0.0116)	(0.0159)	(0.0249)	(0.0492)
Control mean	[0.0482]	[0.0525]	[0.0482]	[0.0525]
English	0.0138	0.0293**	0.0296	0.0901**
	(0.0107)	(0.0141)	(0.0230)	(0.0441)
Control mean	[0.0401]	[0.0440]	[0.0401]	[0.0440]
Math	-0.0010	0.0115	-0.0022	0.0354
	(0.0063)	(0.0085)	(0.0135)	(0.0265)
Control mean	[0.0115]	[0.0165]	[0.0115]	[0.0165]
Science	0.0038	-0.0028	0.0082	-0.0087
	(0.0076)	(0.0113)	(0.0163)	(0.0346)
Control mean	[0.0170]	[0.0207]	[0.0170]	[0.0207]
Social Studies	0.0282**	0.0203	0.0604**	0.0625
	(0.0134)	(0.0179)	(0.0291)	(0.0556)
Control mean	[0.0649]	[0.0711]	[0.0649]	[0.0711]
Remedial courses				
Math	-0.0031	0.0298	-0.0067	0.0916
	(0.0212)	(0.0245)	(0.0454)	(0.0761)
Control mean	[0.2057]	[0.1937]	[0.2057]	[0.1937]
Science	-0.0045	-0.0267	-0.0096	0.1099
	(0.0258)	(0.0320)	(0.0552)	(0.0995)
Control mean	[0.3336]	[0.3316]	[0.3316]	[0.3316]
Observations	6,383	4,579	6,383	4,579

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.389 points for the ELA models and BW = 4.955 points for the Content models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. Columns 1 and 3 present the ITT estimates for ML-ELs whose 9th grade school most common EL placement was ELA only and Columns 2 and 4 presents the ITT estimates for ML-ELS whose 9th grade school most common EL placement was ESL all content. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade.. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: “Math Models” and “Math Strategies. Remedial Science = Integrated Physics and Chemistry. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A28. *Effects of Reclassification on Student Achievement by Most Common G9 EL Program*

	ITT		CATE	
	(1) ESL ELA	(2) ESL All Content	(3) ESL ELA	(4) ESL All Content
STAAR EOC Achievement (SDs)				
Algebra I	-0.0166	-0.0632	-0.0356	-0.1940
	(0.0345)	(0.0462)	(0.0739)	(0.1436)
Control mean	[−0.5351]	[−0.4925]	[−0.5351]	[−0.4925]
Observations	6140	4348	6140	4348
Biology	0.0664*	-0.0468	0.1428*	-0.1443
	(0.0361)	(0.0435)	(0.0780)	(0.1365)
Control mean	[−0.7591]	[−0.7384]	[−0.7591]	[−0.7384]
Observations	6117	4296	6117	4296
English I	0.0017	-0.0139	0.0037	0.0429
	(0.0256)	(0.0328)	(0.0549)	(0.1010)
Control mean	[−0.6167]	[−0.6209]	[−0.6167]	[−0.6167]
Observations	6166	4351	6166	4351
English II	-0.0116	-0.0424	-0.0250	-0.1298
	(0.0300)	(0.0464)	(0.0643)	(0.1429)
Control mean	[−0.7351]	[−0.7351]	[−0.7351]	[−0.7227]
Observations	5856	4161	5856	4161
US History	-0.0207	-0.0151	-0.0442	-0.0490
	(0.0471)	(0.0693)	(0.1006)	(0.2244)
Control mean	[−0.7745]	[−0.7093]	[−0.7745]	[−0.7903]
Observations	4153	2900	4153	2900
Retake EOC				
Algebra I	0.0133	0.0301	0.0285	0.0916
	(0.0232)	(0.0293)	(0.0496)	(0.0893)
Control mean	[0.2927]	[0.2760]	[0.2927]	[0.2760]
Observations	6097	4288	6097	4288
Biology	-0.0155	-0.0117	-0.0334	-0.0360
	(0.0240)	(0.0297)	(0.0518)	(0.0913)
Control mean	[0.2574]	[0.2542]	[0.2574]	[0.2542]
Observations	6054	4204	6054	4204
English I	0.0058	0.0198	-0.0125	0.0600
	(0.0247)	(0.0285)	(0.0358)	(0.0865)
Control mean	[0.7847]	[0.7222]	[0.7847]	[0.7722]
Observations	5928	4236	5928	4236
English II	-0.0371	0.0473	-0.0505	0.1461
	(0.0252)	(0.0300)	(0.0548)	(0.1033)
Control mean	[0.7189]	[0.7069]	[0.7189]	[0.7069]
Observations	5599	4027	5599	4027
US History	0.0067	-0.0018	0.0143	0.0389
	(0.0242)	(0.0273)	(0.0516)	(0.0899)
Control mean	[0.1827]	[0.1760]	[0.1827]	[0.1760]
Observations	4078	2826	4078	2826

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth (BW = 5.13 points, ELA models; BW = 4.81 points, Content models) via Calonico et al. (2014). Cols. 1 and 2 present intent-to-treat (ITT) estimates. Cols. 3 and 4 present complier local average treatment effects (CATEs). Cols. 1 and 3 present estimates for ML-ELs whose G9 school's most common EL placement was ELA only. Cols. 2 and 4 present estimates for ML-ELs whose G9 school's most common EL placement was ESL all content. All models include year fixed effects and covariates for student and G8 school characteristics, and district region. Robust standard errors, in parentheses, clustered at the G8 school level. Control means, in brackets, calculated using students who scored below the STAAR cutoff. Observation numbers within the bandwidth are presented below the Mean row. For this analysis, we further restrict the analytic sample to exclude students who did not take the Algebra I EOC before G9. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A29. *Effects of Reclassification on High School Course Failure by Most Common G9 EL Program*

	ITT		CATE	
	(1)	(2)	(3)	(4)
	ESL ELA	ESL All Content	ESL ELA	ESL All Content
Failed English Class	0.0363	0.0422	0.0777	0.1295
	(0.0281)	(0.0335)	(0.0607)	(0.1027)
Control mean	[0.5345]	[0.5494]	[0.5345]	[0.5494]
Math Class	0.0003	0.0205	0.0006	0.0629
	(0.0271)	(0.0344)	(0.0579)	(0.1058)
Control mean	[0.5945]	[0.6067]	[0.5945]	[0.6067]
Science Class	0.0604**	0.0367	0.1292**	0.1125
	(0.0279)	(0.0330)	(0.0604)	(0.1024)
Control mean	[0.4903]	[0.5114]	[0.4903]	[0.5114]
Social Studies Class	0.0098	0.0605*	0.0210	0.1855*
	(0.0259)	(0.0329)	(0.0556)	(0.1037)
Control mean	[0.4611]	[0.4773]	[0.4611]	[0.4773]
Observations	6,208	4,391	6,208	4,391

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.128 points for the ELA models and BW = 4.806 points for the Content models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (LATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. Columns 1 and 3 present the ITT estimates for ML-Els whose 9th grade school's most common EL placement was ELA only and Columns 2 and 4 presents the ITT estimates for ML-ELS whose 9th grade school's most common EL placement was ESL all content. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. For this analysis, we further restrict the analytic sample to exclude students who did not take the Algebra I EOC before ninth grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

TABLE A30. *Effects of Reclassification on High School Graduation by Most Common G9 EL Program*

	ITT		CATE	
	(1) ESL ELA	(2) ESL All Content	(3) ESL ELA	(4) ESL All Content
Ever Graduated	-0.0271*	-0.0099	-0.0590*	-0.0315
	(0.0146)	(0.0187)	(0.0318)	(0.0596)
Control mean	[0.9074]	[0.9049]	[0.9074]	[0.9049]
Observations	5690	4082	5690	4082
Graduated On-time	-0.0321	-0.0454*	-0.0698	-0.1442*
	(0.0206)	(0.0240)	(0.0405)	(0.0746)
Control mean	[0.8273]	[0.8459]	[0.8273]	[0.8459]
Observations	5690	4082	5690	4082
Graduated College Ready	-0.0138	-0.0608***	-0.0608***	-0.1932***
	(0.0164)	(0.0220)	(0.0358)	(0.0746)
Control mean	[0.1045]	[0.1181]	[0.1045]	[0.1181]
Observations	5690	4082	5690	4082
Drop Out	0.0271*	0.0190	0.0590*	0.0315
	(0.0146)	(0.0189)	(0.0318)	(0.0596)
Control mean	[0.9074]	[0.0951]	[0.9074]	[0.0951]
Observations	5690	4082	5690	4082

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.212 points for the ELA models and BW = 4.854 points for the Content models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. Columns 1 and 3 present the ITT estimates for ML-Els whose 9th grade school's most common EL placement was ELA only and Columns 2 and 4 presents the ITT estimates for ML-ELS whose 9th grade school's most common EL placement was ESL all content. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. On-time' graduation is defined as graduating in spring of the fourth year after entering high school. 'College-ready' is defined as completing 4 English classes, 3 math classes (including algebra I and II), 3 science courses (including biology and chemistry or physics), 3 social studies, and 2 world language classes. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A31. *Effects of Reclassification on Student Course Taking by School ML-EL Concentration*

	ITT		CATE	
	(1) High ML-EL %	(2) Low ML-EL%	(3) High ML-EL %	(4) Low ML- EL %
AP/IB/DE courses				
Any course	0.0183	0.0365*	0.0699*	0.0846*
	(0.0231)	(0.0194)	(0.0421)	(0.0454)
Control mean	[0.1008]	[0.1008]	[0.1008]	[0.1008]
>1 course	0.0282*	0.0252*	0.0506	0.0585
	(0.0159)	(0.0153)	(0.0330)	(0.0360)
Control mean	[0.0506]	[0.0506]	[0.0506]	[0.0506]
English	0.0271*	0.0301**	0.0611**	0.0698**
	(0.0150)	(0.0131)	(0.0286)	(0.0310)
Control mean	[0.0418]	[0.0418]	[0.0418]	[0.0418]
Math	0.0101	0.0088	0.0150	0.0205
	(0.0079)	(0.0081)	(0.0174)	(0.0189)
Control mean	[0.0137]	[0.0137]	[0.0137]	[0.0137]
Science	-0.0008	0.0078	0.0155	0.0182
	(0.0116)	(0.0095)	(0.0202)	(0.0221)
Control mean	[0.0189]	[0.0189]	[0.0189]	[0.0189]
Social Studies	0.0246	0.0217	0.0416	0.0503
	(0.0187)	(0.0173)	(0.0371)	(0.0402)
Control mean	[0.0690]	[0.0690]	[0.0690]	[0.0690]
Remedial courses				
Math	0.0452*	0.0090	0.0063	0.0210
	(0.0243)	(0.0288)	(0.0611)	(0.0669)
Control mean	[0.2020]	[0.2020]	[0.2020]	[0.2020]
Science	-0.0159	0.0122	0.0214	0.0283
	(0.0266)	(0.0347)	(0.0750)	(0.0808)
Control mean	[0.3150]	[0.3150]	[0.3150]	[0.3150]
Observations	4,301	3,608	4,198	3,608

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.622 points for the high ML-EL concentration models and BW = 4.823 points for the lower ML-EL concentration models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. Columns 1 and 3 present the ITT estimates for ML-ELs attending schools in the bottom third of the ML-EL distribution (less than 9% are ML-ELs) and Columns 2 and 4 presents the ITT estimates for ML-ELS attending schools in the top third of the ML-EL distribution (more than 15% of students are ML-ELs). All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. AP = Advanced Placement. IB = International Baccalaureate. DE = Dual Enrollment. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. Remedial Math includes two courses: “Math Models” and “Math Strategies. Remedial Science = Integrated Physics and Chemistry.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A32. *Effects of Reclassification on Student Achievement by School ML-EL Concentration*

	ITT		CATE	
	(1) High ML-EL %	(2) Low ML-EL %	(3) High ML-EL %	(4) Low ML-EL %
STAAR EOC Achievement (SDs)				
Algebra I	-0.1077**	-0.0866**	-0.3263**	-0.2010**
	(0.0516)	(0.0430)	(0.1615)	(0.1017)
Control mean	[-0.4832]	[-0.5332]	[-0.4832]	[-0.5332]
Observations	4145	3505	4145	3505
Biology	0.0280	-0.0340	0.0864	-0.0790
	(0.0533)	(0.0420)	(0.1640)	(0.0979)
Control mean	[-0.7838]	[-0.7153]	[-0.7838]	[-0.7153]
Observations	4108	3484	4108	3484
English I	-0.0047	-0.0456*	-0.0143	-0.1063
	(0.0381)	(0.0275)	(0.1153)	(0.0648)
Control mean	[-0.6102]	[-0.5920]	[-0.6102]	[-0.5920]
Observations	4158	3505	4158	3505
English II	-0.0167	-0.0387	-0.0502	-0.0892
	(0.0417)	(0.0410)	(0.1261)	(0.0947)
Control mean	[-0.7139]	[-0.6959]	[-0.7139]	[-0.6959]
Observations	3937	3360	3937	3360
US History	-0.0677	-0.0175	-0.2169	-0.0386
	(0.0695)	(0.0681)	(0.2244)	(0.1500)
Control mean	[-0.7637]	[-0.7794]	[-0.7637]	[-0.7794]
Observations	2539	2507	2539	2507
Retake EOC				
Algebra I	0.0648**	0.0358	0.1934**	0.0831
	(0.0257)	(0.0338)	(0.0789)	(0.0788)
Control mean	[0.2491]	[0.3180]	[0.2491]	[0.3180]
Observations	4071	3489	4071	3489
Biology	0.0008	0.0316	0.0024	0.0734
	(0.0295)	(0.0294)	(0.0900)	(0.0686)
Control mean	[0.2508]	[0.2513]	[0.2508]	[0.2513]
Observations	4031	3474	4031	3474
English I	-0.0390	0.0411	-0.1221	0.0955
	(0.0325)	(0.0307)	(0.1025)	(0.0718)
Control mean	[0.7721]	[0.7697]	[0.7721]	[0.7697]
Observations	3891	3490	3891	3490
English II	-0.0083	0.0165	-0.0252	0.0380
	(0.0334)	(0.0327)	(0.1016)	(0.0756)
Control mean	[0.6971]	[0.7074]	[0.6971]	[0.7074]
Observations	3686	3321	3686	3321
US History	0.0389	-0.0199	0.1248	-0.0437
	(0.0304)	(0.0306)	(0.0996)	(0.0674)
Control mean	[0.1647]	[0.1836]	[0.1647]	[0.1836]
Observations	2466	2477	2466	2477

Note. Estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth (BW = 5.36 points, high ML-EL % models and BW = 4.81 points, lower ML-EL % models) (Calonico et al., 2014). Cols. 1 and 2 present intent-to-treat (ITT) estimates. Cols. 3 and 4 present complier average treatment effects (CATEs). Cols. 1 and 3 presents estimates for ML-ELS attending schools where > 15% of students are ML-ELs (top third of the distribution). Cols. 2 and 4 present estimates for ML-ELs attending schools where < 9% are ML-ELs (bottom third of the distribution). All models include year fixed effects and covariates for student and G8 school characteristics, and district region. Robust standard errors, in parentheses, clustered at the G8 school level. Control means, in brackets, calculated using students who scored below the cutoff. For this analysis, we exclude students who took the Algebra I EOC before G9. STAAR = State of Texas Assessments of Academic Readiness. EOC = End of course. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A33. *Effects of Reclassification on Student Achievement by School ML-EL Concentration*

	ITT		CATE	
	(1) High ML-EL %	(2) Low ML-EL %	(3) High ML-EL %	(4) Low ML-EL %
Failed English Class	0.0158	0.0448	0.0477	0.1039
	(0.0341)	(0.0363)	(0.1031)	(0.0846)
Control mean	[0.5361]	[0.5224]	[0.5361]	[0.5224]
Math Class	-0.0008	0.0195	-0.0241	0.0454
	(0.0350)	(0.0350)	(0.1055)	(0.0816)
Control mean	[0.5897]	[0.5892]	[0.5897]	[0.5892]
Science Class	-0.0078	0.0472	0.0321	0.1097
	(0.0330)	(0.0349)	(0.0995)	(0.0814)
Control mean	[0.4868]	[0.5047]	[0.4868]	[0.5047]
Social Studies Class	0.0522	0.0213	0.1576	0.0944
	(0.0320)	(0.0345)	(0.1028)	(0.0800)
Control mean	[0.4697]	[0.4467]	[0.4697]	[0.4467]
Observations	4,190	3,543	4,190	3,543

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.356 points for the high ML-EL concentration models and BW = 4.814 points for the lower ML-EL concentration models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (LATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A34. *Effects of Reclassification on High School Graduation by School ML-EL Concentration*

	ITT		CATE	
	(1) High ML-EL %	(2) Low ML-EL %	(3) High ML-EL %	(4) Low ML-EL %
Ever Graduated	-0.0261	-0.0161	-0.0813	-0.0380
	(0.0182)	(0.0180)	(0.0574)	(0.0597)
Control mean	[0.8879]	[0.9261]	[0.8879]	[0.9261]
Observations	3880	3219	3880	3219
Graduated On-time	-0.0454*	-0.0321	-0.1413*	-0.0755
	(0.0243)	(0.0252)	(0.0771)	(0.0972)
Control mean	[0.8515]	[0.8117]	[0.8117]	[0.8117]
Observations	3880	3219	3880	3219
Graduated College-ready	-0.0388**	-0.0159	-0.1207**	-0.0373
	(0.0184)	(0.0245)	(0.0597)	(0.0758)
Control mean	[0.0925]	[0.1281]	[0.0925]	[0.1281]
Observations	3878	3218	3878	3218
Dropped out	0.0261	0.0180	0.0183	0.0243
	(0.0180)	(0.0180)	(0.0574)	(0.0597)
Control mean	[0.1121]	[0.0739]	[0.1121]	[0.0739]
Observations	3880	3219	3880	3219

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 5.622 points for the high ML-EL concentration models and BW = 4.823 points for the lower ML-EL concentration models). These bandwidths were selected via Calonico et al.'s (2014) bandwidth selection procedure for each subgroup. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (LATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed-effects and covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE A35. *Effects of Reclassification on Credit Accumulation*

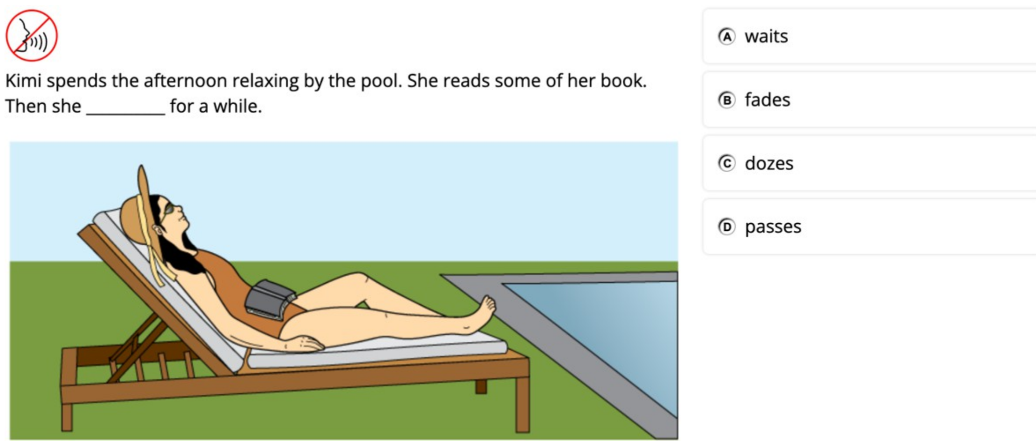
	ITT		CATE	
	(1) Year FE	(2) +Pretreatment Covariates	(3) Year FE	(4) +Pretreatment Covariates
4 English Credits	-0.0626***	-0.0621***	-0.1433***	-0.1418***
	(0.0232)	(0.0233)	(0.0538)	(0.0539)
Control mean	[0.2862]	[0.2862]	[0.2862]	[0.2862]
3 Math Credits	-0.0122	-0.0025	-0.0279	-0.0058
	(0.0285)	(0.0286)	(0.0654)	(0.0653)
Control mean	[0.5586]	[0.5586]	[0.5586]	[0.5586]
3 Science Credits	-0.0188	-0.0147	-0.0429	-0.0335
	(0.0241)	(0.0237)	(0.0553)	(0.0541)
Control mean	[0.2681]	[0.2681]	[0.2681]	[0.2681]
3 Social Studies Credits	0.0102	0.0236	0.0234	0.0538
	(0.0271)	(0.0273)	(0.0621)	(0.0625)
Control mean	[0.6077]	[0.6077]	[0.6077]	[0.6077]
2 World Language Credits	-0.0443	-0.0399	-0.1015	-0.0912
	(0.0297)	(0.0295)	(0.0685)	(0.0678)
Control mean	[0.4389]	[0.4389]	[0.4389]	[0.4389]
Observations	6,071	5,664	6,071	5,664

Note. Cells present point estimates generated by local linear non-parametric regression using a triangular kernel and optimal bandwidth from the first stage regression of a reclassification indicator on the running variable (BW = 4.912859353491786 points). This bandwidth was selected via Calonico et al.'s (2014) bandwidth selection procedure. Columns 1 and 2 present intent-to-treat (ITT) estimates, showing the impact of passing the STAAR threshold for reclassification--conditional on meeting TELPAS requirements--on each outcome. Columns 3 and 4 present complier local average treatment effects (CATEs), second stage results from two-stage least squares estimation that show the impact of being reclassified on outcomes for students induced to reclassify by passing the STAAR threshold. All models include year fixed effects. Columns 2 and 4 also include covariates for student and eighth-grade school characteristics, and district region within Texas to increase precision. Standard errors, presented within parentheses, are heteroskedasticity-robust and clustered at the eighth-grade school level. Control means, in brackets, are calculated using students who scored below the STAAR threshold and were within the bandwidth. The number of observations within the bandwidth is presented below the Mean row. The analytic sample includes students who 1) were first time eighth graders and EL-classified in Texas public schools from 2013-2017; 2) have a valid test score on both TELPAS and STAAR; 3) scored 'advanced high' on the writing, speaking, and listening domains of TELPAS; 4) took the second administration of STAAR; 5) did not receive testing accommodations on the second attempt that would make them ineligible for reclassification after 2014; and 6) re-enrolled in a public school in Texas in 9th grade. TELPAS = Texas English Language Proficiency Assessment System. STAAR = State of Texas Assessments of Academic Readiness. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Appendix B: Figures

FIGURE B1

TELPAS Eighth Grade Reading Sample Item, 2023



STAAR Eighth Grade Sample Item, 2023

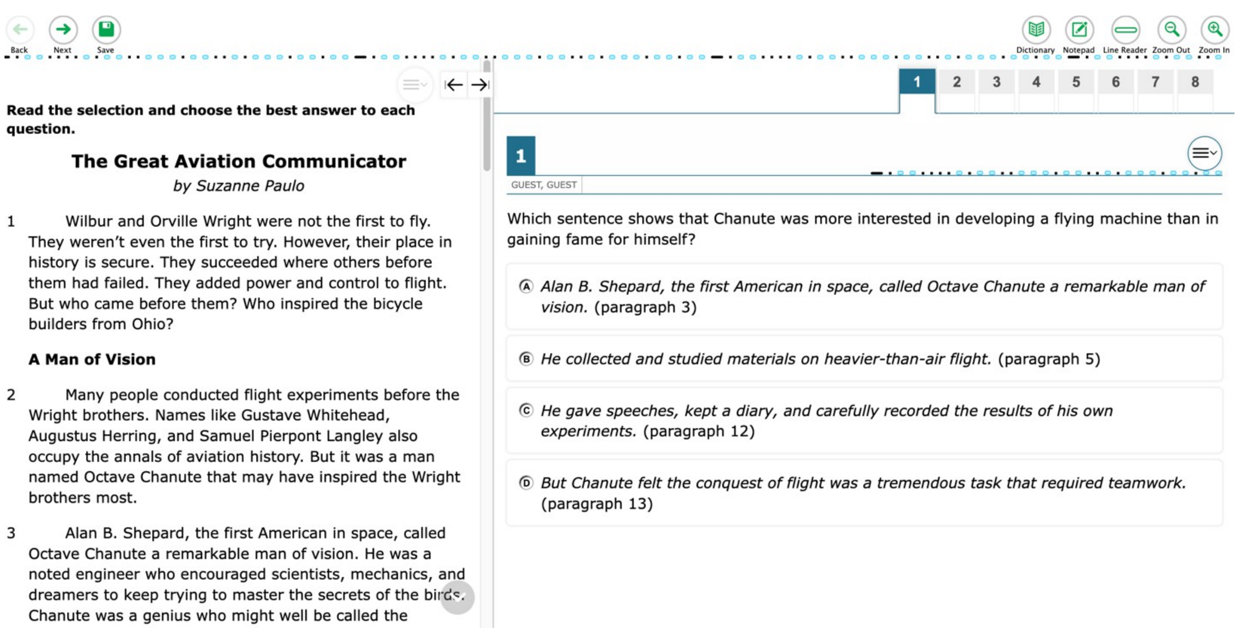
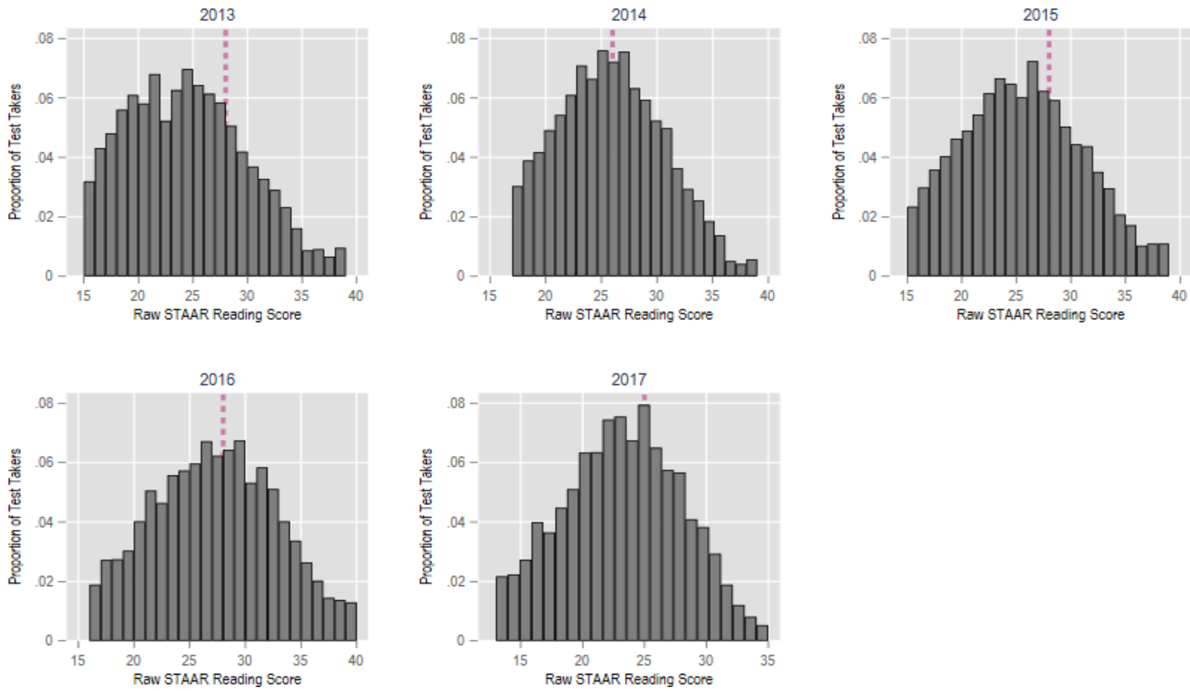
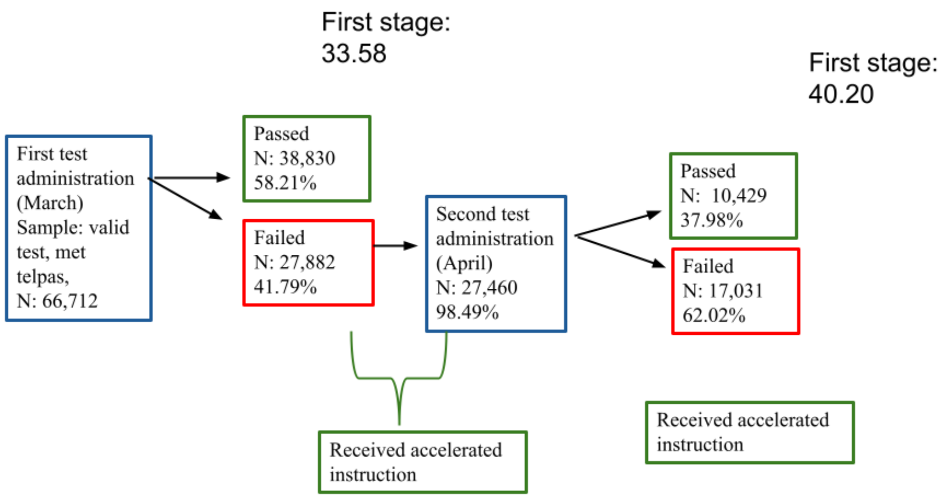


FIGURE B2. ***Distribution of Running Variable by Year***



Note. Graphs show the distribution of raw scores on the STAAR reading assessment for students in the analytic sample by year. STAAR= State of Texas Assessments of Academic Readiness

FIGURE B3. *Percentage of Grade 8 ML-ELs Who Took STAAR Reading Administrations 1 and 2*



Note. The figure shows the percentage of Grade 8 ML-EL students who took, passed, and failed administrations of STAAR Reading at Time 1 and Time 2. The first box on the left corresponds to the sample described in Appendix A Table A2, Column 3, consisting of all students who were EL-classified, had valid test scores for STAAR and TELPAS, and met all TELPAS conditions to be reclassified: achieving "advanced high" on listening, speaking, and writing domains. Of this group, 58.21% passed on their first attempt. Of the 41.79% who failed in T1, 98.49% of them retook the exam during T2. In between T1 and T2, they received state-mandated accelerated instruction. Of T2 test takers, 37.98% pass, and 62.02% fail. The first stage estimation results for both T1 and T2 are shown above the figure. For T1, the first stage is 33.58 percentage points, likely because many students pass on the second administration and are subsequently reclassified. The first stage for T2 is 40.20 percentage points. Given the stronger first stage and cleaner comparison in T2, we selected that administration for our RD analysis.