Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1247

The effects of third-grade retention on multilingual students: A gateway or a gatekeeper?

Diana Quintero
Vanderbilt University

This paper estimates the effect of test-based grade retention on multilingual students classified as English Learners (ML-ELs). This policy could provide an opportunity for ML-ELs to develop English language proficiency and master academic content or put them at increased risk of worse academic and labor market outcomes. I use a regression discontinuity design leveraging Texas's test-based grade retention policy to estimate the effects of this policy on ML-ELs' short- and long-term outcomes. Under Texas policy, students scoring below the reading test cutoff are required to repeat third grade and receive additional reading support. I find that, for students close to the test score cutoff, grade retention and supplemental services have a positive effect on math and reading achievement, as well as increasing the probability of reclassification. These positive effects do not, however, translate into higher post-secondary enrollment or earnings. I find small and not statistically significant effects of grade retention on ML-ELs' probability of enrollment in a post-secondary institution. I also find that earnings for marginally retained students are statistically the same as those for promoted students in young adulthood. Finally, I find suggestive evidence that the effects of retention are larger in schools with higher expenditures on retained students or on bilingual education. This paper underscore the importance of analyzing how education policies designed with all students in mind affect ML-ELs in the short- and long-term.

VERSION: July 2025

Suggested citation: Quintero, Diana. (2025). The effects of third-grade retention on multilingual students: A gateway or a gatekeeper?. (EdWorkingPaper: 25-1247). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/kq9z-yw58

The effects of third-grade retention on multilingual students: A gateway or a gatekeeper?

Diana Quintero*

Vanderbilt University

Abstract

This paper estimates the effect of test-based grade retention on multilingual students classified as English Learners (ML-ELs). This policy could provide an opportunity for ML-ELs to develop English language proficiency and master academic content or put them at increased risk of worse academic and labor market outcomes. I use a regression discontinuity design leveraging Texas's test-based grade retention policy to estimate the effects of this policy on ML-ELs' short- and long-term outcomes. Under Texas policy, students scoring below the reading test cutoff are required to repeat third grade and receive additional reading support. I find that, for students close to the test score cutoff, grade retention and supplemental services have a positive effect on math and reading achievement, as well as increasing the probability of reclassification. These positive effects do not, however, translate into higher post-secondary enrollment or earnings. I find small and not statistically significant effects of grade retention on ML-ELs' probability of enrollment in a post-secondary institution. I also find that earnings for marginally retained students are statistically the same as those for promoted students in young adulthood. Finally, I find suggestive evidence that the effects of retention are larger in schools with higher expenditures on retained students or on bilingual education. This paper underscore the importance of analyzing how education policies designed with all students in mind affect ML-ELs in the short- and long-term.

*I thank the Education Research Center at Houston University for providing the data for this study. I am also grateful to Felipe Barrera-Osorio, Sean Corcoran, Christopher Candelaria, and Ilana Umansky for thoughtful comments, and to Emily Phillips-Galloway, David Figlio, Michael Hansen, Matthew Kraft, Umut Özek, Jon Valant, and seminar participants at the Society for Research on Educational Effectiveness conference, the Association of Public Policy Analysis and Management, and the Peabody Education Policy seminar at Vanderbilt University for their valuable feedback. This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA-NSF Grants Program" from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those AERA or NSF. Any errors are my own.

1 Introduction

Multilingual students classified as English learners (ML-ELs) are an important, often overlooked, subgroup in American education. ML-ELs account for 10.4% of all students nationwide but have historically been provided unequal access to educational resources, contributing to lower academic outcomes (Gándara et al., 2003; NCES, 2023). Further, education policies are often blind to ML-ELs' strengths and unique challenges, typically framing their needs only in the domain of language learning (Gándara and Rumberger, 2009; Ruiz, 1984). However, while developing English language proficiency and literacy skills, ML-ELs simultaneously learn academic content, an experience shaped by education policies. A growing body of literature examines the effect of policies specifically targeting ML-ELs. These policies include classification as English Learners, reclassification as English proficient, and modes of instruction. Yet, little research investigates how broader education policies—those geared toward all students—impact ML-ELs' academic outcomes. Given this landscape, research into the impact of broader education policies on ML-ELs is needed.

In this paper, I respond to this gap by estimating the causal effect of test-based grade retention, which is one broader education policy. I focus on the short- and long-term effects of this policy on ML-ELs in Texas from 2003 to 2009. During this period, Texas required schools to retain third-grade students who did not pass the state reading test, unless a grade placement committee decided otherwise. This test-based retention policy creates a discontinuity in the probability of retention at the test score cutoff used to determine reading passing standards. I can therefore use a regression discontinuity (RD) approach to estimate local causal effects of grade retention. The RD design compares students who scored just below to those who scored just above the threshold for grade promotion. I also use a fuzzy RD to account for noncompliance. The primary outcomes I examine include ML-ELs' academic achievement in elementary school, probability of reclassification as English proficient, high school graduation, post-secondary enrollment, and earnings in young adulthood. Importantly, Texas's retention policy required schools to provide additional instructional support to students who failed the state reading test. Accordingly, my estimates of the impact of retention capture the combined effect of retention and additional reading support.

U.S. schools have a long history of implementing grade retention to improve student achievement. Recently, this policy has gained attention as a strategy to allow students to catch up in the wake of the pandemic (Özek and Mariano, 2023). Further, third-grade retention laws have been included as a policy option to improve early-grade literacy under the science of reading movement. Currently, 26 states plus the District of Columbia require or allow retention for third graders who are not reading proficiently (Redding and Carlo, 2023). Advocates of this educational practice posit that an extra year will provide students with time to catch up academically and acquire foundational skills that are needed for subsequent grades, particularly when retention occurs in early grades (Goos et al., 2021). For ML-ELs, the additional year along with any additional services targeting retained students, may act as a gateway to help them catch up in reading and attain reclassification as English proficient sooner. Attaining reclassification before middle school, when students are exposed to more rigorous course work, often benefits ML-ELs' academic and non-academic trajectories (Chin, 2021; Pope, 2016). Conversely, opponents argue that grade retention could affect students socially as they are detached from their friends and experience feelings of failure, impacting their academic self-concept and self-confidence. As a result, repeaters may disengage from the classroom and drop out at higher rates than promoted students (Anderson et al., 2005; Goos et al., 2021). Because ML-ELs often feel stigmatized and devalued in the classroom and are already at a higher risk of dropping out, retention may function as a gatekeeper for them (Brooks, 2019; Dabach, 2014; Flores et al., 2015; Soland and Sandilos, 2021).

To understand whether grade retention is a gateway or a gatekeeper policy for ML-ELs, this paper answers the following research questions: First, what is the causal effect of third-grade retention on ML-ELs' short- and long-term educational and labor market outcomes, including academic achievement, reclassification as English proficient, high school completion, post-secondary enrollment, and earnings? Second, does the effect of third-grade retention differ across student subgroups or school characteristics?

Using administrative data for Texas third-grade students from school academic years 2002–2003 to 2008–2009, I find that marginally retained ML-ELs who scored just below the cutoff point used to determine eligibility for promotion and received accelerated instruction have higher academic achievement and are more likely to attain reclassification in subsequent grades than marginally promoted students. I find small effects that are indistinguishable from zero of supplemental services and grade retention on students' probability of ever graduating or formally dropping out. However, third-grade retention and supplemental reading instruction increases ML-ELs' probability of graduating within 9 years after completing third grade by 23 percentage points, primarily because third-grade retention reduced the likelihood of being retained again. The large and positive effects on achievement and reclassification do not, however, translate into higher post-secondary enrollment or earnings. More specifically, I find a small and not statistically significant effect of grade retention on ML-ELs' probability of enrollment in a post-secondary institution in Texas 9 to 12 years after promotion from third grade. I also find that earnings for marginally retained ML-ELs are statistically the same as those for promoted students in young adulthood. These results are not sensitive to the inclusion of covariates or to bandwidth choice and are not driven by differential attrition between retained and promoted students.

In sum, the results show that third-grade retention and accelerated instruction may act as a gateway for ML-ELs in the short term; however, when post-secondary enrollment and earnings are considered, grade retention and supplemental services do not benefit (or harm) ML-ELs. These findings are aligned with other research suggesting that short-run effects of education policies, including Head Start, class size, school choice, and teacher unions, can differ from their longer-run effects (Beuermann and Jackson, 2020; Chetty et al., 2011; Deming et al., 2014; Lovenheim and Willén, 2019; Ludwig and Miller, 2007). Accordingly, this paper underscores the importance of examining the long-term effects of broader education policies on ML-ELs.

Further, the subgroup analysis by student and school characteristics shows interesting patterns. First, the effects of grade retention on academic achievement, graduation, and earnings are more pronounced for students who were seven or eight years old in the third grade than for older thid-grade students. Second, the effects of retention on high school completion are higher for students who took the third-grade reading test in Spanish in comparison to those who took the test in English, indicating the importance of assessments in students' home language. Third, the effects of grade retention on reclassification outcomes and on-time graduation are higher among schools with high per-pupil expenditures and expenditures on bilingual programs. Finally, schools with more resources for retained students seem to have higher effects of retention on reclassification, high school graduation, and earnings. Though these estimations have less statistical power, the results suggest the importance of school resources for retained ML-ELs.

The results of this paper make three important contributions to the literature. First, this paper provides evidence of causal effects of grade retention for ML-ELs. Although the effects of test-based retention have been studied in multiple contexts and across grades (Eren et al., 2017; Greene and Winters, 2007; Jacob and Lefgren, 2009), the impact on ML-ELs has been largely understudied, with one exception (Figlio and Özek, 2020). As ML-ELs are retained at higher rates than non-EL students (while 14.4% of third-graders are ML-ELs, 21% of third-grade retained students are ML-ELs) and face unique learning and testing challenges, it is critical to assess the efficacy of this policy for ML-ELs specifically (U.S. Department of Education, 2016). Second, I investigate the overlooked role of school characteristics, including resources for extra support and bilingual education, in moderating the effects of grade retention.

Third, I provide evidence of the impact of third-grade retention on different short- and long-term outcomes. Most research estimates the impact of third-grade retention on academic achievement and behavioral outcomes up to five years after retention (Greene and Winters, 2007; Hwang and Koedel, 2023; Jacob and Lefgren, 2004; Mumma and Winters, 2023). Only two studies provide evidence of the effect of retention on high school completion (Mariano et al., 2024; Schwerdt et al., 2017). My paper contributes to this scant literature on the effects of third-grade retention on high school completion. Additionally, to my knowledge, this paper is the first one to estimate the impact of early grade retention on post-secondary enrollment and earnings in young adulthood in the US. Estimating short- and long-term outcomes is fundamental to expanding our understanding of the dynamic of education policies over time.

The paper proceeds as follows: In Section 2, I summarize the results of prior literature estimating the effects of grade retention. Then, Section 3 presents the conceptual framework used to explain how grade retention could act as a gatekeeper or gateway for students. Section 4 describes the Texas retention policy. In Section 5, I describe the data and the outcomes of interest. Section 6 presents the identification strategy and evidence supporting its validity. Section 7 shows the effects of grade retention on student academic achievement, reclassification, high school completion, post-secondary entry, and earnings. Section 8 presents the heterogeneous effects of retention. Section 9 shows robustness checks, and Section 10 concludes.

2 Literature Review

Early descriptive, empirical research on grade retention suggests that retained students have lower academic achievement and higher high school dropout rates in comparison to promoted students (Jimerson, 2001; Shepard and Smith, 2017). Estimates from these studies are likely biased given the non-random nature of grade retention and the presence of unobserved variables associated with retention and student outcomes. Later studies have relied on the expansion of test-based grade retention policies over time and regression discontinuity (RD) designs that compare outcomes for students on either side of a score threshold for grade retention, providing local (around the cutoff) causal effects of grade retention.

These causal studies found that test-based retention is more beneficial for students in early grades and mostly for the first years after retention. Studies from Florida, Chicago, Mississippi, and Indiana showed that third-grade retention along with supporting services improves student achievement. Scholars estimating same-age effects found that retained students do better in reading and math than promoted students 1 or 2 years after retention (Greene and Winters, 2007; Jacob and Lefgren, 2004).[1]

Similarly, same-grade comparisons showed that retained students largely outperform their socially promoted peers in reading through middle school, though the achievement gains decrease over time (Hwang and Koedel, 2023; Mumma and Winters, 2023; Schwerdt et al., 2017). For example, in Indiana, marginally retained students scored 0.513 standard deviations higher on reading than their marginally promoted peers. This effect on achievement decreased to 0.20 SD and was no longer significant by seventh grade (Hwang and Koedel, 2023). Further, the initial increase in student achievement does not seem to increase student graduation. Schwerdt et al. (2017) and Mariano et al. (2024) found that grade retention in elementary school had no impact on the likelihood of high school graduation in Florida and in New York City, respectively.

[1]Same-age estimates compare retained students to promoted students in the following years after retention, when students are at the same age but in different grades. Same-grade estimates compare retained and promoted students when both groups have reached the same grade but are different ages.

The effect of retention on older students has been less studied, and results have varied across different cities. For example, sixth-grade retention had no effect on either math or reading achievement 2 years later or on the likelihood of high school completion in Chicago (Jacob and Lefgren, 2004,0). Conversely, in New York City, retention among fifth graders improved students' reading scores in the two subsequent grades but resulted in fewer high school credits accumulated and a higher probability of dropping out by Grade 12 (Mariano and Martorell, 2013; Mariano et al., 2024).

In contrast, the evidence suggests that retaining students in eighth grade has a negative effect on students' academic trajectories. Eighth-grade retention appears to decrease the probability of high school completion and college entry but has null effects on college graduation (Jacob and Lefgren, 2009; Larsen and Valant, 2023). Studies that examine non-academic outcomes found the effect of eighth-grade retention changes over time. The Louisiana eighth-grade retention policy, for example, reduced the likelihood of juvenile crime (Eren et al., 2017) but increased the probability of violent crime conviction by age 25 (Eren et al., 2022). The mixed effects of grade retention on crime in Louisiana highlight the need to estimate its impact on long-term outcomes, including earnings, which are an important goal of education and are critical to students' long-term well-being.

Only one study has rigorously examined test-based grade retention's effects specifically for ML-ELs. Using Florida's third-grade retention policy, Figlio and Özek (2020) found beneficial effects on ML-ELs' academic trajectories. The authors reported that third-grade ML-ELs who were retained and received additional instructional support reduced their time to be classified as English Proficient by half, were less likely to take remedial English courses, and were more likely to take advanced courses and college-credit-bearing courses in high school than similar ML-ELs who were promoted.

The current paper advances Figlio and Özek's (2020) study in two ways. First, it provides evidence of grade retention on ML-ELs' high school completion, postsecondary access, and earnings. Considering labor market outcomes is particularly important as they represent the opportunity cost of an extra year in school and, thus, help to assess the policy's cost-effectiveness. Second, this paper provides evidence of retention in another context, Texas, where ML-ELs can take the test in

Spanish, receive bilingual education, and are provided with different types of support during the retained year. While Texas's retention policy requires schools to provide "accelerated instruction" which should be delivered at least once a week and in small groups, Florida's legislation requires retained students to receive 90 minutes of daily reading instruction and be assigned to high-quality teachers. For external validity concerns, it is important to consider the effects of retention in different contexts.

3 Grade Retention as a Gateway or Gatekeeper for ML-ELs

The opportunity gap framework is useful to understand both how ML-ELs are more susceptible to test-based retention policies and how this policy may affect their outcomes differently from non-ELs. In this framework, Milner (2012) shifts the onus of differences in student performance away from individual deficiencies and highlights the fewer learning opportunities available for nonwhite students.

In the case of ML-ELs, opportunity gaps emerge from differences in school resources, instruction, assessments, and teachers. ML-ELs typically enroll either in urban high-minority schools or in suburban schools where they comprise a small portion of the student body. Neither of these schooling contexts provides adequate opportunities for ML-ELs to succeed. Urban schools are typically under-resourced, serve disproportionately economically disadvantaged students, and are more likely to be staffed with less experienced teachers (Gándara et al., 2003; Quintero and Hansen, 2021) while suburban schools offer fewer linguistic support services to meet the needs of ML-ELs (Dondero and Muller, 2012). Further, research has shown that ML-ELs are socially and academically isolated, receive less instructional time in content areas, tend to be placed in low-track classes, and have teachers who display lower expectations of them (Callahan, 2005; Katz, 1999; Murphy and Torff, 2019; Olsen, 1997; Umansky, 2016; Valdes, 2004). Additionally, ML-ELs are often assessed on content knowledge across subjects using tests written in English. These tests are constructed for native speakers, and thus test results may not accurately reflect ML-ELs' knowledge and skills (Abedi, 2006; Gàndara and Baca, 2008; Robinson-Cimpian et al., 2016). Together, these factors help to explain the low performance of ML-ELs on standardized tests and their higher likelihood of being retained.

It is plausible that test-based grade retention policies offer a learning opportunity for ML-ELs to master both the English language and academic content. For example, retained students could be assigned to a high-quality teacher and receive academic support during the retained year. These policies can act as gateways to additional instructional time and teacher quality, addressing the opportunity gaps mentioned by Milner (2012) including teacher quality and access to rigorous curriculum. I hypothesize that retention could be a gateway policy if retained students have achievement gains that persist over time, improving graduation rates. Further, according to human capital theory, if grade retention improves students' learning and productivity, it would also improve students' earnings (Goldhaber and Özek, 2019; Weiss, 1995).

Conversely, grade retention could compound existing gaps in opportunities. Retention could reinforce teachers' low expectations (Garcia et al., 2019; Murphy and Torff, 2019; Ochoa, 2013). These beliefs could translate into access to a less rigorous curriculum, emphasis on teaching to the test, or misdirection to special education. These practices could lead to higher dropout rates. Additionally, because retained students reach the legal age to drop out sooner than promoted students, they may accrue less grades of attainment. Dropping out may lead to lower enrollment in post-secondary education and lower earnings.

Additionally, opportunity gaps may vary across schools, and therefore the ways in which grade retention widens or worsens these gaps may vary as well. Schools with more resources could provide better support to retained ML-ELs, but for those enrolled in under-resourced schools, the extra year may act instead as a hindrance as students delay their entry into the workforce. I make an important contribution by conducting a subgroup analysis based on school characteristics to understand the potential role of school resources in moderating the effect of retention.

4 Texas's Grade Retention Policy

The 76th Texas Legislature enacted the Student Success Initiative (SSI) in 1999 to ensure all students receive academic support in mathematics and reading so that they can succeed through high school. The SSI has a grade advancement requirement for third-, fifth-, and eighth-grade students tied to performance on the state standardized tests. The legislation came into effect for third-grade students in the 2002–2003 school year, for fifth-grade students in the 2004–2005 school year, and for eighth-grade students in the 2007–2008 school year (TEC §28.0211, 1999).

From 2002–2003 to 2011–2012, Texas used the Texas Assessment of Knowledge and Skills (TAKS) for retention decisions. Specifically, third-grade students had to pass the state standard in the reading section of the state content-area TAKS test to be promoted to fourth grade, and students in Grades 5 and 8 had to achieve passing standards in both the reading and mathematics sections of the TAKS to be promoted to sixth and ninth grade, respectively. The passing standard was set by the Texas State Board of Education, and the tests were graded by external raters. These tests were offered in English and Spanish for students in elementary grades. The legislation allowed students to take each assessment up to three times. Students who failed the first administration had to receive accelerated instruction in the subject they failed before they could take the test a second time.

A Grade Placement Committee (GPC) was established to guide the instructional support for those who did not demonstrate proficiency in the second administration (TEC §28.0211, 2010). The committee was composed of the principal or principal's designee, the student's teacher(s) of the subject in which the student failed to demonstrate proficiency, and the student's parent or guardian. The GPC also decided whether students should retake the test in the third administration. Students who failed the third administration were required to be retained unless the GPC unanimously decided the student was very likely to perform well in the next grade level with additional instructional support (e.g., participate in accelerated instruction). For ML-ELs, the retention decision was also made in agreement with the Language Proficiency Assessment Committee (LPAC). The data suggests that the GPC and LPAC committees intervened on a regular basis to promote students when they fail the third administration of the test.

Along with grade advancement requirements, the Texas Education Agency (TEA) created grant programs to provide districts with additional resources for accelerated instruction of students who demonstrated difficulty in reading. The amount of funding awarded to school districts was based on the number of students who failed TAKS. Individual schools had flexibility over different aspects of the accelerated program, but the Texas Education Code required schools to provide accelerated instruction to students at least once a week. Also, by mandate, classes had a maximum student-to-teacher ratio of 10:1 and were taught by a certified teacher in the subject of instruction (reading or math). However, each school decided when the program would be delivered (i.e., during school hours or after school hours operations), how many times per week, and the curriculum to follow. For example, Austin Independent School District provided three sessions of accelerated instruction in 2003–2004: 1) from January to March (15 weeks), 2) in April between test administrations, and 3) a summer session. All sessions were conducted in small groups (five to eight students per teacher) and for three hours per week. The curriculum focused on decoding skills and comprehension skills aligned to the TAKS reading test (Curry, 2004).

5 Data

This study uses data from the Texas Education Agency (TEA), the Texas Higher Education Coordinating Board (THECB), and the Texas Workforce Commission (TWC). These data sources are linked by the Texas Education Center. The TEA data include information on student demographics, graduation, English Learner classification, program participation, standardized test scores, and English language proficiency scores on TELPAS (Texas English Language Proficiency Assessment System). These data also include school and district-level information about school characteristics and district expenditure data on expenses related to the SSI, and especially to accelerated instruction. The THECB data provides information on student participation in post-secondary institutions in Texas. The TWC data includes information on earnings for industries covered by unemployment insurance in Texas. The sample for this study includes all students who took TAKS in third grade from 2002–2003 through 2008–2009. This period allows for longer-term investigation of student outcomes of interest (e.g., high school and college graduation and earnings).

Tables 1 and 2 present student and school characteristics of first-time third graders between 2003 and 2009, by ML-EL status. Table 1 shows important differences between ML-ELs and their native English speaker peers. First, ML-ELs were more likely to be in poverty than non-ELs. Second, the share of students who were retained was higher among ML-ELs than among non-ELs. While 4.5% of ML-ELs were retained in third grade, only 2.1% of non-ELs were retained. Additionally, ML-ELs exhibited lower levels of math and reading achievement than their non-EL peers.

Table 2 shows that ML-ELs and non-ELs attended different types of schools. ML-ELs tended to enroll in schools with a high concentration of economically disadvantaged students, Hispanic/Latino students, and other ML-ELs. Further, student achievement in math and reading was lower in schools where the typical ML-EL attended. ML-ELs also attended schools with slightly lower per-pupil expenditures and expenditures for helping struggling students (accelerated instruction and SSI). This table shows important differences in student demographics and school characteristics between ML-ELs and non-ELs that may have contributed to higher retention rates among ML-ELs and to different effects of the grade retention policy for these students.

I restricted the full sample in several ways for the analysis. First, I excluded students with disabilities because they are assessed through a modified version of the test (7.6%). Second, I dropped students who did not re-enroll in a public school in the state of Texas in the fall following the third administration of third-grade TAKS (3.8%). This restriction was necessary to determine retention status. Finally, I limited the sample to students who have a valid TAKS score since my research design relies on observing those scores.[2] With these limitations, the analytic sample is comprised of 475,126 students representing 86.4% of all ML-EL students in third grade during the study period. This sample is demographically comparable to the full sample of all ML-ELs in third grade.

An important feature of Texas's grade retention policy was that it required students to take the reading test up to three times (March, April, and June) and receive accelerated instruction between each test administration if they fail. As a result, I could not include all students taking the test at any point because students around the threshold could have differed both in the amount of accelerated instruction received and the number of test attempts. Instead, a valid RD analysis requires focusing on a particular test administration. However, using the first or second administration provided smaller jumps in the probability of retention around the cutoff given that students who failed the test could retake it and be promoted. Indeed, the probability of retention decreased to only 2.9 and 12.7 percentage points around the threshold for students who took the first and second test administration, respectively. This smaller jumps would have led to less efficient

[2] Some students have scores on the test, but the scores have a flag indicating the student was absent or had a waiver.

estimates. Because the retention decision was based on the third attempt, this analysis focuses on students who failed the reading test on the first two attempts and took the summer test. The high percentage of students retaking the test and having valid scores suggests that students did not self-select to retake the test (See Figure 1).[3]

Using data on only students who retook the test in the summer could limit the generalizability of the results if students who took the test in March were systematically different from students who took the test in the summer. Table 3 compares the characteristics of students across the three test administrations. Students who retook the test in April and in the summer were more likely to be male, live in economically disadvantaged households, and be Hispanic/Latino. Not surprisingly, retakers had lower academic achievement. These students were more likely to have been retained prior to third grade and to have lower reading scores and less likely to meet the standard for math.[4] Additionally, retakers were less likely to be reclassified as English proficient by the end of third grade and had lower scores in the four domains of English language proficiency: Reading, Listening, Speaking, and Writing.

In sum, students who took the summer test were in a more vulnerable situation than students who took the first test; these differences may have contributed to lower academic outcomes for retakers in the future (i.e., reclassification, graduation, college entry). Indeed, students who retook the test for the third time had lower academic achievement, graduation rates, and post-secondary enrollment, on average. These differences were driven by the fact that students who passed the first attempt of the test had considerably higher subsequent outcomes than students who did not pass the test until their third attempt.

Given the difference in demographic and academic achievement between students who only took the test one time and those who took it on the third attempt, the effect of narrowly failing the

[3]Among the 14,938 students who should have taken the test for a second time, 79% were LEP-Exempt (because they had been in US schools for less than a year) in test 1 which is probably why they did not have to retake the test. Among the students who failed the previous two administrations, 9.5% were LEP-Exempt in the second test, 22% of them had a parental or committee waiver, 8.8% were determined by the ARD committee to be inappropriate for TAKS, and 43.57% were absent. The average reading score in the second attempt was 16.7 points for students who took the third test while the average for students who didn't take the third test was 14.8 points.

[4]Though test-based grade retention under SSI is for grades 3, 5, and 8, students in other grades can still be retained.

test could have differed for these two populations of students. Students who retook the test may have required even more instructional time than those who took the test only one time; thus, a full extra year could have provide them with the knowledge required to catch up and progress in school. Conversely, retakers may also have had lower perceptions of themselves, and the additional year could have harmed their self-esteem even more. Consequently, they comprise an important group of students to inform policy and practice.

5.1 Outcomes

I estimate the effect of grade retention on a variety of indicators of academic and labor market success: (a) academic achievement; (b) reclassification; (c) high school completion; (d) enrollment in post-secondary institutions; and (e) average quarterly earnings. Below, I describe the measure used for each outcome.

Academic Achievement

I estimate the effect of third-grade retention on students' achievement in math and reading in the fourth and fifth grades. To do so, I first standardize ML-ELs' TAKS raw scores across all third-grade ML-ELs within subject, grade, and year, so that the mean of each test score is zero and the standard deviation equals 1. The raw points are the number of questions a student answered correctly on the test. Then, I compare the standardized scores of marginally retained students to those of promoted students when both groups had reached fourth and fifth grade. Because marginally retained students would have been older and would have been exposed to more schooling than marginally promoted students when they took any grade-level test after retention, this same-grade comparison will identify the effect of retention only in the absence of age effects and time-in-school effects (Schwerdt et al., 2017).

An alternative way to estimate the effects of grade retention is to compare marginally retained students to promoted students in the years after retention (same-age comparisons). Same-age comparisons show the effect of the policy in the years immediately following retention; however, the estimates compare students who had been exposed to a different curriculum and took a different test (third-grade test vs. fourth-grade test). In Texas, fourth graders take a test that has more questions and a greater difficulty than the third-grade test. A lower percentage of fourth-grade students achieve higher scores, suggesting that the expected learning rate for fourth graders is higher than for third graders. Therefore, same-age comparisons may be confounded by differences in tests and the average rate of learning across grades (Schwerdt et al., 2017). A vertically aligned test would help to mitigate these concerns, but the reading and math tests in Texas were not vertically aligned during the time of this study.

Reclassification

To estimate the effect of grade retention on reclassification, I use the following outcomes: the number of years students were classified as EL since their first time in third grade, the probability of reclassification in the years and grades after retention, and the probability of reclassification at any grade during their time at school. Although existing research finds mixed results of the effects of reclassification on student achievement and graduation, some evidence suggests that reclassification improves ML-ELs' academic achievement, particularly when reclassification occurs in elementary grades (Chin, 2021; Pope, 2016). Given that EL status is consequential for students, it is important to estimate whether grade retention accelerates or hinders their readiness for reclassification.

High School Graduation and Dropout

The third set of outcomes includes high school completion, graduation on time, and dropout. High school completion is defined as whether students had received a standard high school degree by the end of 2022. At that point, students of the 2003 cohort had had 19 years to complete high school and were on average 27 years old, while students of the 2009 cohort had had 13 years to complete high school and were on average 21.2 years old. In Texas, students can graduate from high school under two academic plans: 22 credit and 26 credit plans. Since the latter plan provides students with more content classes, I also estimate the effect of retention on students' likelihood of graduating under the 26-credit plan. While graduating from high school has important economic implications for students (Heckman et al., 2014), graduating on time provides students with an earlier entry to college and more years in the labor market. I measure on-time graduation as a binary indicator for whether the student graduated by the spring of the ninth year after being promoted from third grade. For instance, for first-time third graders in the 2003-2004 cohort who were not retained in third grade, on-time graduation was by the spring of 2013, and for those who were retained, on-time graduation was by 2014. Students who were still enrolled in school after 9 years or who dropped out of high school were coded as "nongraduates". Finally, dropout is defined as whether a student in the administrative data is formally identified as a dropout by the end of the 2022 year and does not have a graduation date. Students identified as dropouts but re-enrolling in schools and having a graduation date are considered high school completers. Students who were not in the graduation, dropout, or exit files were excluded from the analysis of graduation and dropout because these students may have moved to a private school or completed high school outside Texas (about 6% of students in the sample were excluded).

Post-Secondary Enrollment

I link student-level data from TEA to the THECB data to measure postsecondary enrollment. About 33% of ML-ELs in the sample were enrolled in any post-secondary institution (i.e., 2-year public institution, 4-year public institution, for-profit institution, or private institution) between 2009 and 2021.[5] Students may be absent from the higher education dataset either because they worked in Texas and indeed did not attend higher education or because they moved out of state. I use the workforce dataset (TWC) to distinguish between these scenarios. Specifically, for each year of data, I assume that students without information on higher education enrollment for that year but who had income in that year were indeed not attending a post-secondary institution in Texas. Students without either earnings or higher education information were not included in the analysis (11%).

I create two variables for post-secondary enrollment. The first variable is a binary indicator equal to 1 if students are ever observed enrolling in a higher education institution between their 9th and 12th year after satisfactorily completing third grade. This measure allows both promoted and retained students from the most recent cohort, 2008–2009, to have at least 3 years of potential post-secondary enrollment. The second variable equals 1 if students are ever observed enrolling in higher education when they are 17 to 20 years old. While enrollment by years post-third grade completion provides the same number of years for retained and promoted students to enroll in higher education,

[5]This percent is 35% for non-ELs who took the summer reading test.

the variable for enrollment at ages 17 to 20 accounts for the fact that retained students spend an additional year in K-12 education. I also create a separate indicator for enrollment in community colleges.

Labor Market Outcomes

An important contribution of this paper is estimating the effects of retention on earnings. This outcome is highly important as it allows us to estimate the long-term effects of retention and consider the opportunity cost of being in school for an additional year. I link Unemployment Insurance (UI) records from the Texas Workforce Commission on student employment and earnings to student-level data from TEA. UI records include only taxable reported earnings for non-federal employment within the state that was eligible for unemployment benefits. Measures of students' earnings and employment are available until the second quarter of 2024, which means that all cohorts have at least 5 years of income information after their expected high school exit. Matches between the student records and UI were identified for 77% of third-grade students.[6] Similar to post-secondary enrollment, if a student is not in the workforce dataset it may be because they did not have income, worked in the informal sector in Texas, or moved out of state. I assume that students enrolled in a school or a post-secondary institution in Texas in a given year and are not in the Texas workforce dataset had zero earnings in that year. This assumption may exclude earnings such as federal work, seasonal work that may go underreported to unemployment insurance, and self-contract employment, including work in the gig economy. Yet, it is not a major limitation given that research suggests that workers with no or partial coverage in UI are about 10% of US employment (Hotz and Scholz, 2002) and that administrative data are less likely to miss self-reported earnings than survey data (Abraham et al., 2017).

The TWC data were used to generate quarterly measurements of students' earnings and the number of quarters worked in a year. I estimate the effect of grade retention on students' average quarterly earnings (measured in 2022 dollars and transformed into logarithms) between the ages of 20 and 22. This time frame provides students with two to three years to complete higher education

[6]To compare, 91% of third-grade students not classified as EL have information on wages. The lower match rate among ML-ELs could be due to these students moving out of state, leaving the country, or working in the informal sector.

and two years of income after higher education. I also estimate the effects on earnings for students who were not enrolled in a higher education institution at ages 20 to 22 as a proxy for full-time employment given that the data do not include hours worked.

6 Methods

To estimate the effect of grade retention on ML-ELs' outcomes, I use an RD analysis that exploits Texas's use of test thresholds for grade retention decisions. This approach allows me to identify causal estimates in the local area around the thresholds. Because students are not randomly selected for grade retention, simple comparisons of averages in outcomes between students who were promoted and those who were retained will not provide causal estimates of the grade retention effect. Systematic differences between these two groups that I do not observe and are correlated with retention, such as student motivation, or maturity, would likely bias the estimates. However, it can be argued that students scoring within a narrow range around the grade retention cutoff are on average equivalent, mimicking a randomized experiment (Imbens and Lemieux, 2008; Lee and Lemieux, 2010).

For each outcome, I estimate the following equation:

$$Y_{i,t} = \beta_0 + \beta_1 below_{i,t} + f(score_{i,t}) + \beta_2 below_{i,t} * f(score_{i,t}) + \epsilon_{i,t}, \tag{1}$$

where $Below_{i,t}$ equals 1 if student i, who was for the first time in third-grade in year t scored below the minimum threshold. The forcing variable, $score_{i,t}$, is student i score on the third attempt of the reading test TAKS. This score is in raw points; I recenter the score so that zero is the passing score. $f(.)$ is a function of the score. In the model, β_1 provides the local intent-to-treat (ITT) estimates for the effect of being retained and receiving accelerated instruction in the following year. My outcome variables, described in more detail in the data section, include academic achievement, reclassification, high school completion, post-secondary enrollment, and earnings.

I use non-parametric models (local linear projections) with an optimal bandwidth calculated using the Calonico et al. (2014) bandwidth selection procedure. I provide sensitivity checks within several bandwidths that are approximately 0.8 and 1.2 times the optimal bandwidth (Chin, 2021; Robinson-Cimpian et al., 2016). Standard errors are clustered at the school level. All estimations use a triangular kernel.

Because the GPC along with the LPAC had discretion in retention decisions, students who failed to attain the grade promotion threshold may still have been promoted to fourth grade, resulting in imperfect compliance. I also estimate treatment-on-treated (TOT^c) estimates among compliers using a fuzzy RD to account for noncompliance. In the fuzzy RD, the threshold for retention is used as an instrumental variable for treatment status (retention). The first stage regression uses Equation 1 where the dependent variable is whether the student was retained. The second stage uses the predicted probability of retention to examine the effect on the outcomes of interest. The TOT^c estimates are the ITT estimates upweighted by the compliance rate.

6.1 Identifying Assumptions

To yield valid estimates, an RD approach assumes that the means of the expected potential outcomes are continuous at the cutoff. I present two analyses that suggest the data are very likely to meet this assumption. First, I test for continuity on observable characteristics around the cut score by running the RD models with student demographic characteristics as outcomes. The estimation uses different bandwidths that fall in the range of the optimal bandwidth used in the estimation of each outcome of interest. Figure 2 shows the result of these estimations across different bandwidths. I do not find evidence of discontinuity around the threshold for retention on gender, socioeconomic status, race, age, English language proficiency, and time in the US. Although I cannot test for discontinuity on unobservables, evidence of continuity in observable demographic characteristics reduces concerns of discontinuous changes in the composition of students at the threshold that would lead to biased results.

The second check explores whether the retention threshold was determined exogenously and whether students (or teachers) were able to sort into treatment. Qualitatively, score manipulation was unlikely because tests were graded by an external testing center and students were very unlikely to predict the number of correct answers associated with scoring close to the cutoff. Figure 3 presents the distribution of the reading test scores on the third attempt; visually it seems that there is no manipulation. I test for disproportionate bunching of students on either side of the cutoff using the Cattaneo et al. (2018) density test. The result suggests that I fail to reject the null hypothesis ($p>0.05$) and therefore there is no statistical evidence of systematic manipulation of the running variable.[7]

Three additional assumptions are needed for the TOT^c estimates to show the average causal effect of retention at the cutoff among compliers (students whose retention status changes as their scores are just slightly lower or higher than the cutoff). First, scoring below the threshold must actually exert an effect on the likelihood of retention. Figure 4 plots the probability of third-grade retention as a function of students' scores on the third attempt of the third-grade TAKS reading assessment. This figure presents evidence that the probability of retention changes discontinuously at the cutoff. Specifically, the estimated probability of third-grade retention decreases by 32.4 percentage points around the threshold. It is not surprising that the probability does not jump from 0 to 1 given the GPC and LPAC's roles in retention decisions.

Second, scoring below the threshold must only affect the outcomes by changing the probability of grade retention. This exclusion restriction assumption is plausible because the results of the summer test were used mainly for retention decisions. However, students' scores on the TAKS were also used along with students' English Language proficiency and teachers' subjective evaluations to determine reclassification status. Indeed, students who passed the third attempt of the reading test were 6.5 percentage points more likely to be reclassified as English Proficient at the end of third grade. This discontinuity around the threshold implies that the effect of retention could be confounded with the effect of reclassification. However, only 1.7% of students scoring below the threshold for promotion were reclassified. This small percentage of the sample may not impose a threat to the validity of the estimates, but my estimates show the combined effect of being retained and maintaining EL status for at least a year. In a supplementary analysis, I excluded the small number of students who were reclassified in the sample (around 5%), and the estimates were very similar. The last assumption for valid TOT^c estimates is that test scores should affect the probability of retention only in one direction, which is evident in Figure 4.

[7]An analysis of manipulation by year yields the same conclusion: there is no evidence of manipulation in each year ($p > 0.25$).

Though students who scored below the cutoff score for retention were more likely to be retained than students above the cutoff, there were students who failed the test and were promoted (if all students who failed the test had been promoted, then the probability of retention around the cutoff would have decreased from 1 to 0 in Figure 4). The non-compliance rate observed in Figure 4 warrants an examination of systematic differences between compliers (students who failed the test and were retained) and noncompliers (those who failed the test and were promoted to fourth grade). To investigate this further, I regress the probability of compliance on a set of student characteristics including, race, gender, socio-economic status, age, previous retention, number of years in US schools, math scores, and language of the reading test. I also include cohort- and school-fixed effects. The results of this exercise show that compliers were very similar to non-compliers in terms of gender, economic status, and race/ethnicity. However, being older, being retained in first or second grade, meeting math standards, and taking the reading test in English, rather than in Spanish, were characteristics associated with a lower probability of retention conditional on failing the test (See Figure A.1). I also examine compliance rates at the school level. School compliance was positively associated with the percentage of economically disadvantaged students and the percentage of students who met the standard for reading on the first attempt. These analyses suggest that TOT^c estimators apply for younger students, with lower math achievement.

7 Results

Tables 4 to 8 present the effects of third-grade retention on ML-EL's short- and long-term outcomes. All tables have the same format. Column 1 shows the ITT estimates, and Column 2 shows the TOT^c estimates among compliers. The tables also include the optimal bandwidth, the mean of the control group, and the number of observations for each outcome. Figures 5 and 6 provide visual illustrations of the treatment effects. All estimations include the following covariates: gender, economically disadvantaged status, age in third grade, number of years in EL status until third grade, participation in a bilingual program in third grade, first-grade retention, second-grade retention, math achievement in third grade (whether the student met the standard), speaking Spanish at home, taking the reading test in English (students can take it in English or Spanish), school enrollment, student composition (percent of economically disadvantaged students, percent of ML-ELs), region, and cohort-fixed effects. Tables A.1 to A.5 present the results including only cohort-fixed effects —the results are very similar. The ITT estimates show the average local effect of barely failing the threshold for promotion, and the TOT^c estimates show the average causal effect of retention at the cutoff among compliers. Recall that SSI's policy required students who failed the test to participate in the accelerated instruction programs regardless of their retention status. The estimates therefore show the combined effect of failing the test and receiving supplemental support (ITT) and of retention and supplemental support (TOT^c).

Academic Achievement

Table 4 shows the primary RD estimates of the effect of third-grade retention on ML-ELs' academic achievement in Grades 4 and 5. Overall, I find positive and large effects of grade retention and accelerated instruction on students' math and reading achievement in the subsequent grades after retention. Students who narrowly failed the test score 0.33 SD higher in reading and 0.31 SD higher in math in Grade 4 than students who narrowly passed the test (Column 1). Given that one standard deviation in the reading and math test among ML-ELs in fourth grade is 6.92 raw points and 7.15 raw points, respectively, the ITT effect is equivalent to scoring 2.3 and 2.2 additional correct answers on the reading and math tests, respectively. The TOT^c estimates, in Column 2, show the average effect of grade retention among compliers. I find that students who were marginally retained score 0.94 SD higher in reading and 0.93 SD higher in math than their peers who were marginally promoted. These effects are equivalent to having approximately 6.6 additional correct answers on each test, respectively. The effect of retention and accelerated instruction is smaller for writing. Students who were marginally retained score 0.5 SD higher than their peers who were promoted.

By Grade 5, the effect of narrowly failing the third-grade reading test decreases for both subjects. Students who marginally failed the test score 0.25 SD higher in reading and math than their peers who just passed the test. The effect of marginally being retained is also attenuated. Marginally retained students have reading gains of 0.76 SD (5.3 raw points) and math gains of 0.78 SD (5.8 raw points) in comparison to marginally promoted students. Interestingly, the effects of retention and accelerated instruction are very similar across subjects, which could be due to students receiving accelerated instruction in both subjects.

Reclassification

Table 5 presents the combined effects of accelerated instruction and narrowly failing the test (Column 1) or third-grade retention (Column 2) on reclassification as English proficient. I examine several outcomes: the number of years a student was classified as EL, the probability of ever being reclassified, and the probability of reclassification in the following years or grades after retention. I find that narrowly failing the test reduces total years in EL status by 0.25. The TOT^c effects are larger with marginally retained students spending 0.64 fewer years as EL compared to marginally promoted students. The reduction in years in EL status is explained by the effect of narrowly failing the third-grade test on the probability of reclassification as English proficient. ML-ELs who narrowly failed the test are 3.3 and 5.9 percentage points more likely to be reclassified by Grades 4 and 5, respectively, than students who narrowly passed the test (Panel II). The TOT^c estimates show that students who were marginally retained in third grade are 13 percentage points more likely to be reclassified by fourth grade than their promoted peers. By fifth grade, retention in third grade increases the probability of reclassification by 18 percentage points. The increase in the effect of reclassification over time is because this variable is cumulative, so fifth-grade reclassified students include those who were reclassified in third and fourth grade. The effects of narrowly failing the test or being reclassified remain constant after sixth grade, the ITT estimate for attaining reclassification by seventh grade is 4.6 percentage points, very similar to the ITT effect by 11th grade (4.8 percentage points) (See Table A.6).

Findings from same-age comparisons show a slightly different pattern (Table 5, Panel III). The effect of narrowly failing the third-grade reading test on reclassification is statistically insignificant in the year after retention. Students who were retained in third grade are equally likely to be reclassified as promoted students. This effect does not persist over time, however. Students who narrowly failed the third-grade reading test are 3.3 percentage points more likely to attain reclassification 2 years after retention than their peers who passed the test (Column 1). And, being marginally retained increases the probability of reclassification to 10 percentage points two years after retention (Column 2). Failing the test decreases the probability of reclassification in third grade by 6 percentage points; therefore, failing the third-grade test has an immediate negative effect on reclassification, a statistically insignificant effect one year later, and a positive effect 2 years after retention.

High School Graduation

Table 6 shows the effects of accelerated instruction and grade retention on ML-ELs' probability of high school graduation, graduation under a 26-credit plan, dropout, graduation within 4 years in high school, and on-time graduation conditional on being retained (9 years after completing third grade). The findings indicate that grade retention has small effects on ever graduating and on formally dropping out of school, but the estimates are not statistically different from zero. This result is similar to Schwerdt et al. (2017) findings in Florida. Though students who narrowly failed the test are 4.5 percentage points more likely to graduate with a 26-credit plan, the effect of retention on this outcome is imprecise. The effect of grade-retention on on-time graduation conditional on retention (within 9 years after completing third grade) is positive and statistically significant. Students who narrowly failed the test are 12 percentage points more likely to graduate on time (Column 1). Accelerated instruction and marginal retention increases the probability of graduation within 9 years by 22 percentage points (Column 2). This effect represents a 40% increase in on-time graduation.

A potential contributor to the effect of retention on on-time graduation but not on ever graduating is that third-grade retention is associated with a lower probability of retention in subsequent grades. Students who narrowly passed the test are 12.6 percentage points more likely to be retained in Grades 4 through 8. Indeed, if the measure for on-time graduation provides two extra years for promoted students (which means 11 years after completing third grade), then the effect of retention on graduation is no longer statistically significant. Further, both marginally retained and promoted students graduate within 4 years in high school at the same rate. Third-grade retention prevents future retention, given that retained students are more likely to meet both the math and reading standards in Grade 5, which were the criteria for fifth-grade promotion. It is also possible that teachers were reluctant to retain students more than once.

Post-Secondary Enrollment

Table 7 shows the effect of accelerated instruction and grade retention on the probability of enrolling in any type of post-secondary institution. Panel I shows the point estimates of the probability of enrollment in higher education when students were 17 to 20 years old. The effects of third-grade retention on these outcomes are small and statistically insignificant. However, the confidence intervals for the ITT estimates allow me to rule out treatment effects that are less than -2.2 percentage points and greater than 3.2 percentage points for enrollment in any post-secondary institution. Similarly, I can rule out TOT^c effects less than -7.5 percentage points and greater than 8.3 percentage points.

Panel II presents the effects of grade retention and accelerated instruction on post-secondary enrollment 9 to 12 years after students completed third grade. These effects are also small and statistically equal to zero. Marginally failing the test does not affect students' probability of enrollment in type of post-secondary institution, including community colleges in Texas. The confidence intervals for the effects of retention on enrollment 9 to 12 years post-third-grade completion are very similar to the effects by age.

Two hypotheses could explain the statistically insignificant effects on post-secondary enrollment. First, ML-ELs who took the third attempt of the test had very low attainment rates (just a third of students enrolled in post-secondary institutions) therefore increasing enrollment for these students requires additional interventions and addressing other structural barriers that prevent them from having access to post-secondary education. Second, college attendance is relatively insensitive to test score gains, particularly for students with low academic achievement (Backes et al., 2024), thus the positive effect of retention on student achievement on marginally retained students does not improve their college attendance.

Average Earnings

Table 8 shows the ITT and TOT^c effects on students' log average quarterly earnings (expressed in 2022 dollars) when they reached 20 to 22 years old. Students who narrowly failed the test earn on average 1.4% more than students who narrowly passed the test by ages 20 to 22. Yet, this effect is imprecise. Because ML-ELs could have been in school at these ages, and presumably worked fewer hours, the second estimates show the effect of grade retention for ML-ELs who were not enrolled in a post-secondary institution at ages 20 to 22. For these ML-ELs, the effect of marginally being retained is negative (-3.9%) but statistically significant equal to zero. The confidence interval falls between negative 14 and positive 6%; thus, I can rule out that failing the test has a negative impact greater than 14% on earnings or a positive impact greater than 6%. The TOT^c estimates are also negative and imprecise. In a supplementary analysis that examines further if the effect on earnings is due to difference in experience, I find that marginally retained students accrued the same quarters of experience as marginally promoted students between the ages of 15 and 22.

For several reasons, these results on earnings should be interpreted with care. First, I lack information on hours worked, and it may be that one group of students may be more likely to work part-time than the other. Second, I observe students a few years after high school where the returns to education are not as strong as in later years. Third, I do not observe non-cognitive outcomes such as self-belonging, that are also correlated with earnings (Deming, 2017). Fourth, the three most recent cohorts were affected by the COVID-19 pandemic for one or more of the years of the analysis on earnings. For example, students who were 8 years old in 2007 reached the age of 21 by 2020. If the economic effects of the pandemic on the labor market differentially affected retained and promoted students, then the results are confounded by this economic shock.

8 Heterogeneous Effects

I conduct subgroup analyses by student characteristics to examine if the effects of accelerated instruction and grade retention vary across student characteristics. I run ITT models by each of the following student characteristics: gender, age at third grade (less than 8 years old and 9 or more years old), third-grade math score (whether the student scored above or below the median), and language of the third-grade reading test (English or Spanish). Figures 7 to 10 show the results of these estimations along with their 95% confidence intervals. Though the subgroup analysis has less power to statistically detect heterogeneous effects, there are important patterns to note. In Figure 7, we observe that the effect of narrowly failing the reading test on Grade 4 achievement is very similar for boys and girls, yet the decrease in effects by Grade 5 is only among boys. Second, just failing the reading test has more pronounced effects on academic achievement for students who were 8 years old or younger in third grade. For example, younger students who just failed the reading test scored

0.37 SD higher in math in Grade 4 in comparison to their same-age peers who just passed the test. In contrast, the effects on academic achievement among 9-years-old and older students are imprecise (about 25% of students are 9 years old or older in the sample). Third, students whose third-grade math scores were below the median among this population of students experience greater academic effects of retention in Grade 4, but the effects decrease by Grade 5.

Now, turning to the reclassification outcomes (Figure 8), we also observe that the effects of narrowly failing the test are more pronounced for students who were 8 years old or younger in third grade, those whose math scores were below the median, and those who had not been retained before. Figure 9 shows the effect of narrowly failing the reading test on high school completion. The effect of narrowly failing the test on graduation and graduation on a 26-credit plan is positive for younger students and for students who took the test in Spanish. The effects of failing the third-grade test are very imprecise for enrollment in higher education and earnings. There is some suggestive evidence that failing the test has a negative effect on earnings for students who were 8 years old or younger in third grade.

The opportunity gap framework discussed in this paper posits that differences in student performance are driven by differences in learning opportunities for students. Schools play an important role in the opportunities provided to students. If schools improve the instructional opportunities for retained ML-ELs during the additional year, the effects of grade retention may be higher in those contexts. I lack information on the types of supports that schools offered to retained students; however, I have information on school expenditures on different services that serves as a proxy. It is plausible that the positive results of retention along with supplemental services are a byproduct of additional resources for retained students.

To explore this hypothesis, I conduct a subgroup analysis by school characteristics. I divide schools into two groups (those below the median and those at or above the median) for each of the following characteristics: proportion of retained students (in schools where retention is spare the stigma of retention may be worse and therefore the outcomes), SSI expenditures per student who failed the TAKS test in Grades 3 to 8 (recall that TEA provided school districts with non-competitive grants for accelerated instruction under the SSI), per-pupil spending in bilingual education, and total per-pupil spending.[8] For example, schools labeled as having "low per-pupil exp" had expenditures below the median across all schools in the RD sample for each year, while those in the "high expenditures" category had expenditures at or above the median.[9]

Figure 11 presents the effects of narrowly failing the reading test on students' academic achievement in Grades 4 and 5. Schools with high retention rates tend to have greater positive effects than schools with low retention rates. The effect of just failing the test on academic achievement is similar in schools with low and high SSI expenditures. Figure 12 shows that the effects of marginally failing the test on reclassification are greater in schools with high levels of SSI expenditures, total expenditures, and expenditures in bilingual programs. Conversely, the effects of narrowly failing the test on reclassification are not statistically significant for students in schools where these financial resources were below the median.

In Figure 13, we observe the effects of narrowly failing the reading test on graduation outcomes. The effects on graduation rates are more pronounced in schools with high expenditures in SSI, expenditures in bilingual programs, and total expenditures. The effects by school resources are not statistically significant in the long-term, except for earnings, as presented in Figure 14. Students who attended schools with higher SSI expenditures and marginally failed the test, have higher earnings than their peers who were promoted in those high SSI schools.

Though the estimations in the subgroup analysis are imprecise, this analysis provides suggestive evidence that school resources moderate the effects of failing the test, particularly in short-term outcomes. Therefore, the positive effects of retention are likely driven by the accelerated instruction students received and other school services such as bilingual education and teachers.

[8] Though resources were given based on the number of students who failed the reading or math test in a district the variation in this variable mostly comes from SSI funds and not students who failed the tests. The correlation between SSI funds and number of students who failed the test in a school is small ranging from 0.14 to 0.3 in each year. Additionally, schools with high SSI spending include schools with low and high numbers of students who failed the tests.

[9] I calculate the median for each type of spending per year. The median for SSI expenditures ranged from $50-$100, for spending in bilingual education ranged from $1000 to $1700, and for total per-pupil expenditures ranged from $6000 to $6600. All financial resource measures are expressed in 2009 dollars.

9 Robustness Checks

First, I examine the sensitivity of my findings to alternative bandwidths around the test score cutoff. Recall that I use the optimal bandwidth using the Calonico et al. (2014) bandwidth selection procedure for all model specifications. These optimal bandwidths generally range from 3.55 to 4 points on each side of the threshold for retention for the ITT estimation and from 5 to 6 points around the threshold for the TOT^c estimates. In Figures A.2 to A.6 in the Appendix, I present estimates of all my models using bandwidths of 3, 4, 5, 6, 7, and 8 points. This range includes bandwidths more conservative than optimal and encompasses the full range of optimal bandwidths. Additionally, it allows me to compare ITT and TOT^c estimates within the same bandwidth. The figures confirm that my results are not sensitive to bandwidth choice.

Second, examining long-term outcomes poses the challenge that not all third-grade students were observed through high school, post-secondary, and earnings outcomes. Attrition for high school outcomes was small (6%), but it increased to 30% for earnings. Differential attrition might have led to biased estimates if retained students left the sample at higher rates than their promoted peers. To examine this possibility, I estimate a series of RD models in which the dependent variable is whether a student has non-missing information for an outcome or grade. Figure A.7 shows no differential attrition between students who barely failed the test and students who barely passed the test. I also run all the models with the most conservative sample size, and the results are very similar (See Table A.7).

10 Discussion and Conclusions

Educational policies designed with all students in mind often fail to address the unique needs of ML-ELs. Therefore, it is critical to assess the efficacy of these policies on ML-ELs' short- and long-term outcomes. This paper estimates the effect of test-based grade retention, a policy that could either provide an opportunity for ML-ELs to develop English language proficiency and master academic content or put them at an increased risk of worse academic and labor market outcomes. I use a regression discontinuity approach that leverages Texas's grade retention policy to estimate the effect of third-grade retention on ML-EL's academic outcomes, post-secondary enrollment, and earnings.

I find large and positive effects of third-grade retention on student's math and reading achievement in subsequent grades which is consistent with findings from prior studies for non-ELs (Hwang and Koedel, 2023; Schwerdt et al., 2017). I also find a decrease over time in the effects of retention on academic outcomes. Similar to Figlio and Özek (2020), I find ML-ELs who narrowly failed the third-grade reading test improved their reading achievement in the following grades. My ITT estimates, however, are larger than those reported by Figlio and Özek (2020), which could be because in Texas all students who failed the test received supplemental instruction while in Florida the supplemental services were provided only to retained students. In terms of reclassification outcomes, I find that accelerated instruction and narrowly failing the third-grade reading test has a positive effect on students' time to exiting EL status. Students who narrowly missed the threshold for promotion are between 3.3 and 6 percentage points more likely to be reclassified in Grades 4 and 5. These students also spend 0.25 fewer years as EL in comparison to students who narrowly passed the threshold for promotion.

Similar to Schwerdt et al. (2017) and Mariano et al. (2024), I find that third-grade retention has no effect on the likelihood of ever graduating from high school or formally dropping out from school. Yet, I do find that marginally retained students are 23 percentage points more likely to graduate on time (within 9 years after completing third grade) from high school than marginally promoted students. This effect is mostly driven by third grade retention preventing students from being retained in subsequent grades.

The positive effects of grade retention on short-term outcomes do not translate to better academic outcomes after high school. Specifically, marginally retained students enroll in any type of post-secondary institution, including community colleges, at the same rate as their promoted peers. One might expect that the increase in achievement and on-time high school graduation would lead to a higher probability of post-secondary enrollment. However, there are several hypotheses that explain the non statistically significant effect of retention on higher education enrollment. First, ML-EL students face significant barriers to enrolling in higher education, including financial constraints and lack of information. Grade retention and its accompanying additional services are unlikely to be enough to overcome these barriers. Second, academic achievement is not a strong predictor of college enrollment for low-performing students (Backes et al., 2024). Other measures, including school attendance, are more indicative of enrollment in higher education among low-performing students. Third, it is plausible that retention did not affect students' access to rigorous courses in high school which would have prepared them for college. I find marginal evidence that students who were retained graduated under a 26 credit plan. Future research examining the effect of third-grade retention on ML-ELs' attendance and course-taking patterns is needed to improve our understanding of the effect of grade retention on student post-secondary enrollment.

I also find that retention has a negative effect on earnings, though these results are imprecise and statistically insignificant. In the subgroup analysis, I find suggestive evidence that narrowly failing the test decreases earnings by 19% among ML-ELs who were 8 years old or younger in third grade. These effects on earnings, however, should be interpreted with care. First, I lack information on hours worked, which could contribute to differences in earnings. Second, I report earnings in young adulthood when the returns of education may be smaller than in future years. Finally, the results on earnings could be confounded by any effects of the pandemic on the economic labor market.

Overall, the findings show that grade retention along with supplemental services act as a gateway to improved achievement and English proficiency for ML-ELs. Yet, I also find lower earnings in early adulthood for retained students who were 8 years old in third grade, which suggests that retention policies may function as a gatekeeper in the long-term. The difference in the direction of the effects between the short- and long-term could be due to retention effects on short-term outcomes that I do not observe. For example, grade retention may affect students' attendance, self-esteem, and other soft skills, factors that are correlated with labor market outcomes (Deming, 2017). Additionally, grade retention may be an example of an education policy that varies in impact on students' outcomes over time (Beuermann and Jackson, 2020; Chetty et al., 2011; Deming et al., 2014; Lovenheim and Willén, 2019; Ludwig and Miller, 2007).

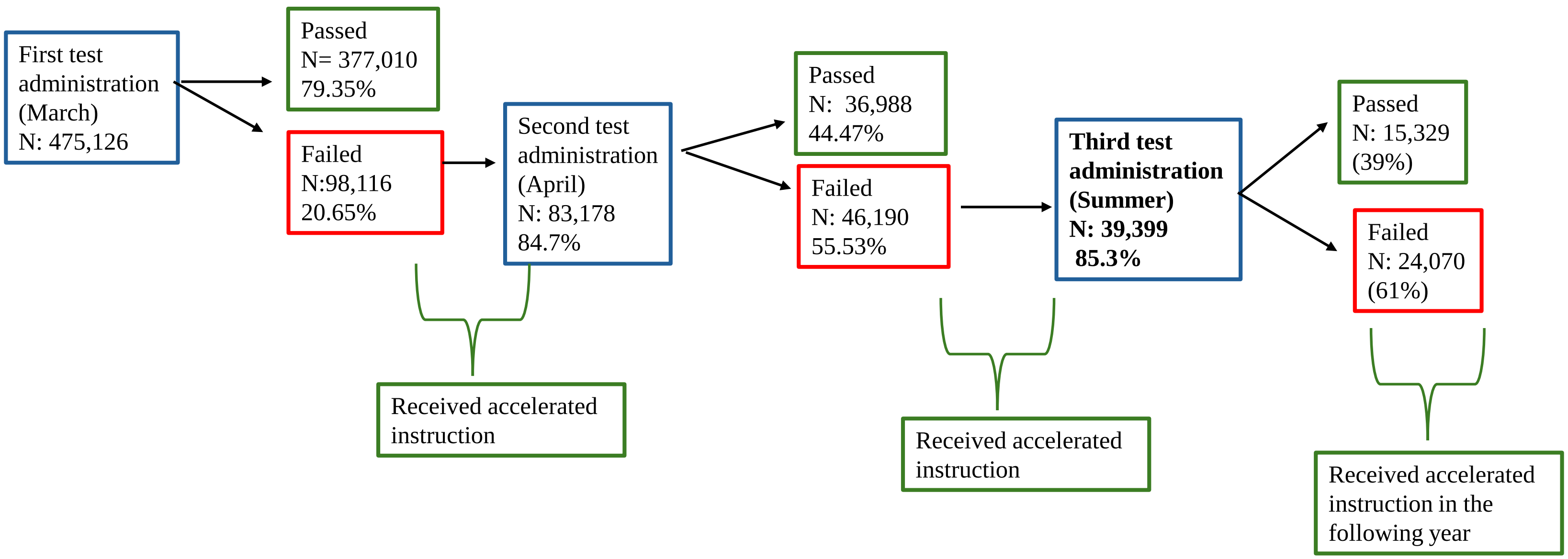
There are important limitations of this paper that need to be acknowledged. First, as with any RD estimation, the external validity of the results is limited to the population around the cutoff. In this study, the sample around the cutoff of the third attempt of the reading test is comprised of ML-EL students who came from very disadvantaged backgrounds, were more likely to be Hispanic/Latinos, and had low English language proficiency. These students also took the reading test three times and received accelerated instruction between each test administration. Accordingly, these students had very low reading proficiency. Additionally, the large TOT^c estimates are among compliers, students whose retention status was determined by their test score. This caveat is important for schools and districts to consider when making retention decisions. The positive results in the short-term for a specific subgroup of students who received extra instruction support does not mean that retention should be applied broadly. Instead, schools could provide accelerated instruction to students who narrowly passed the third attempt of the test. Perhaps these students would have benefited from targeted support over time as they continued with the challenge of learning English and academic content.

Second, the results show a combined effect of grade retention, EL status for at least a year, and receipt of some type of accelerated instruction. The available data for this study do not include information on how the instructional support for students who failed the test was implemented during the year of retention. This lack of information prevented me from estimating the effects of grade retention net accelerated instruction. However, by estimating the effects of retention by the amount of resources devoted to students who failed the test, I find suggestive evidence that SSI resources play an important role in reclassification, graduation from high school, and earnings. Additionally, the effects of retention and accelerated instruction on some outcomes tend to be larger in schools offering bilingual education which typically have bilingual teachers and supports tailored to ML-ELs.

Finally, due to data availability, this paper focuses on estimating the effects of grade retention on only academic outcomes in the years after retention. Retention could negatively affect students' self-esteem and other non-cognitive outcomes that I do not observe, but that are very important for students' educational achievement and well-being. Further research to understand the effect of retention on these outcomes is needed to inform policy.

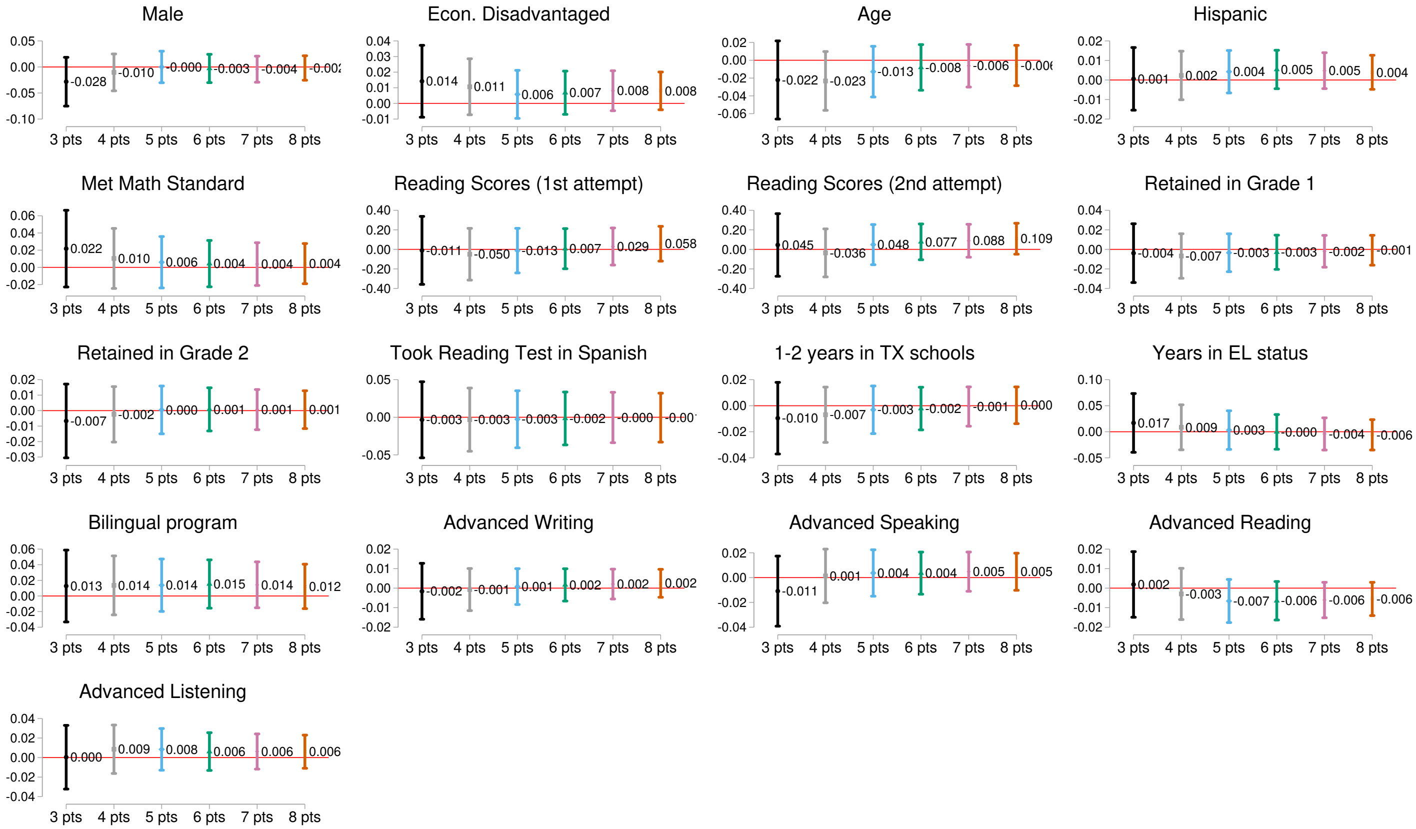
The findings of this paper suggest that ML-ELs with weak academic skills benefit in the short-term from additional instructional time due to supplemental instruction and an extra year in school; however, the gains of retention do not persist over time. The effects of the policy are not statistically significant for other outcomes of interest, including post-secondary enrollment and earnings. Given that retention is costly for schools and could harm students' non-academic outcomes, the policy may not be cost-effective. Therefore, it is important to consider if investing the resources of an extra year in other policies that provide additional instructional time including summer programs and after school programs would improve ML-ELs' long-term academic outcomes.

Figure 1: Number of ML-ELs Who Took the Reading Test on Each Administration



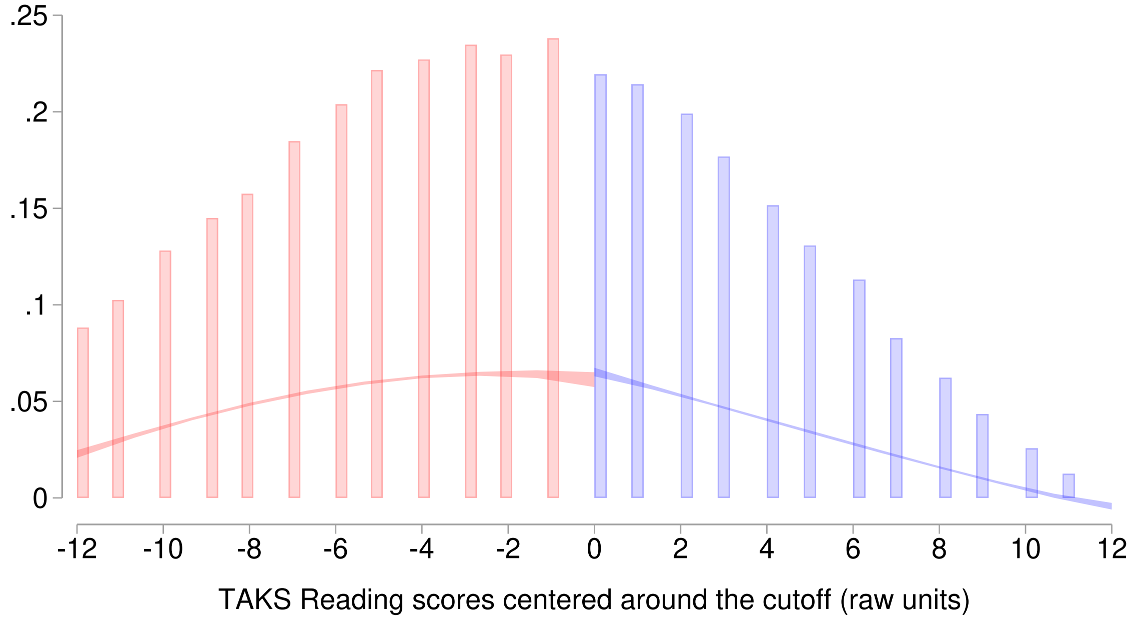
Note: Figure 1 presents the number of ML-ELs who took each administration of the reading test. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 2: Point Estimates at the Cutoff for Retention among Third-Grade ML-ELs



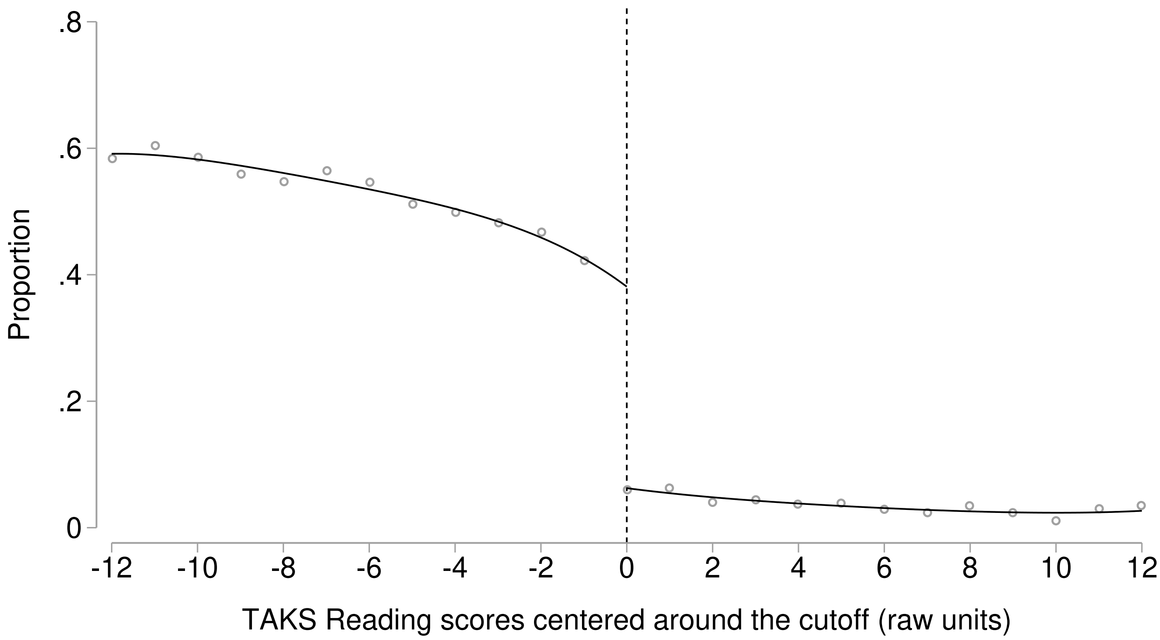
Note: Figure 2 presents the point estimates of non-parametric estimations where the outcome is the covariate shown in the title, and the running variable is students' centered scores on the thrid-attempt of the reading test. The models are estimated for bandwidths of 3, 4, 5, 6, 7, and 8 points around the cutoff for retention. Standard errors are clustered at the school level. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 3: Distribution of Reading Test Scores (Third-Attempt) for Third-Grade ML-ELs



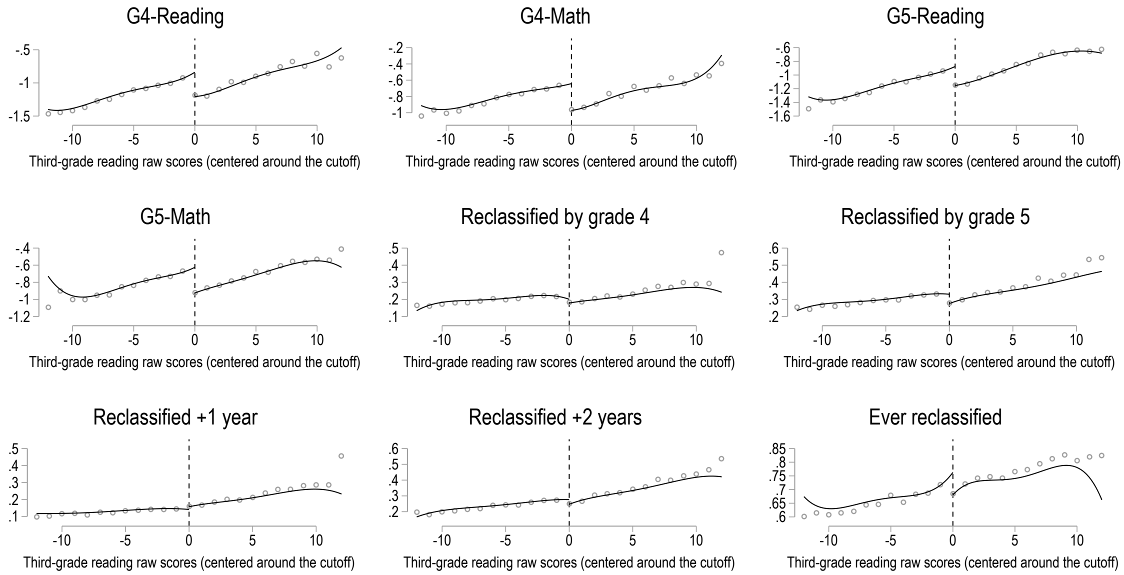
Note: Figure 3 presents the distribution of ML-ELs' reading scores on the third administration of the TAKS test. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 4: Probability of ML-ELs' Third-Grade Retention as a Function of Students' Scores on the Third Attempt of the Reading Test



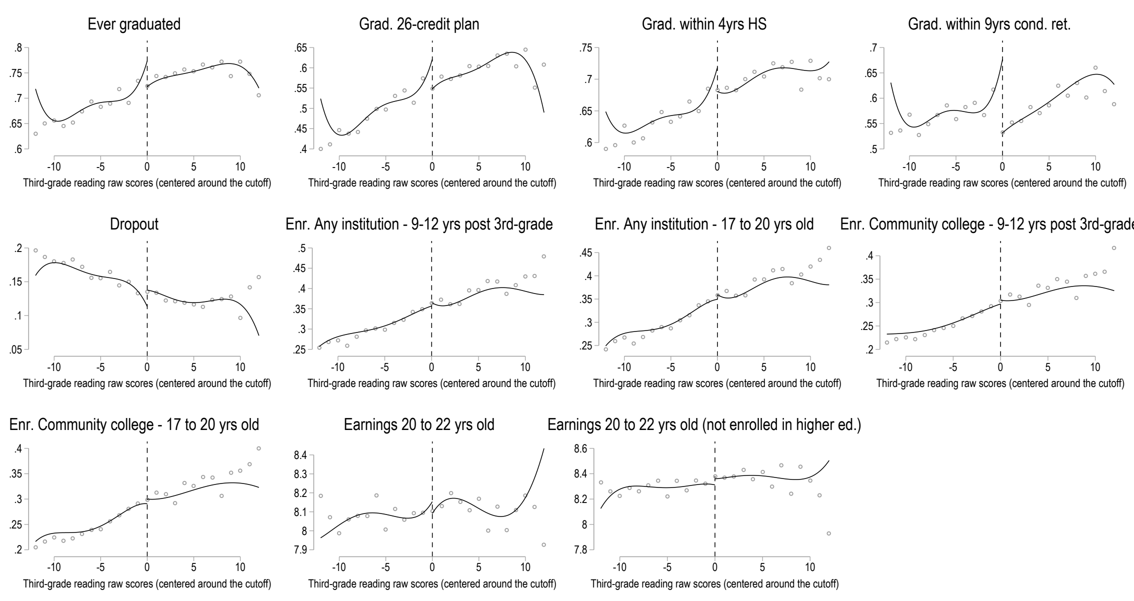
Note: Figure 4 presents a non-parametric estimation of the probability of third-grade retention as a function of students reading scores on the third attempt. Scores are centered at the cutoff. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 5: Effects of Narrowly Failing the Reading Test on Academic Achievement and Reclassification



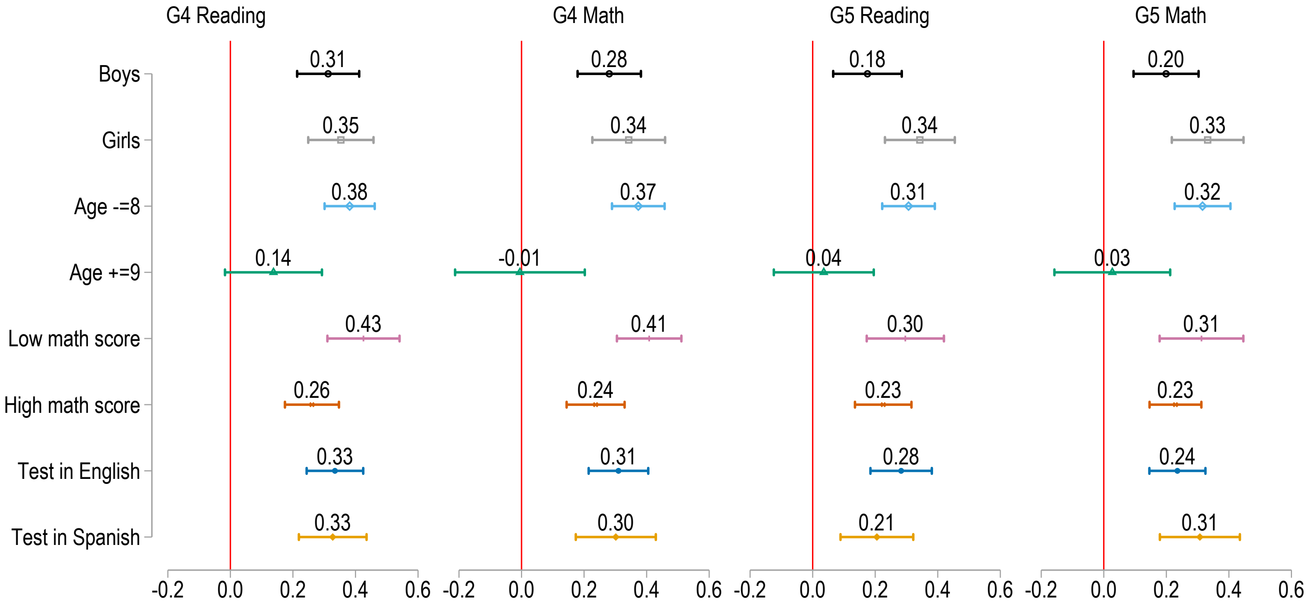
Note: Figure 5 presents non-parametric estimations of the effect of of narrowly failing the reading test on each outcome. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 6: Effects of Narrowly Failing the Reading Test on Graduation, Post-secondary Enrollment, and Earnings



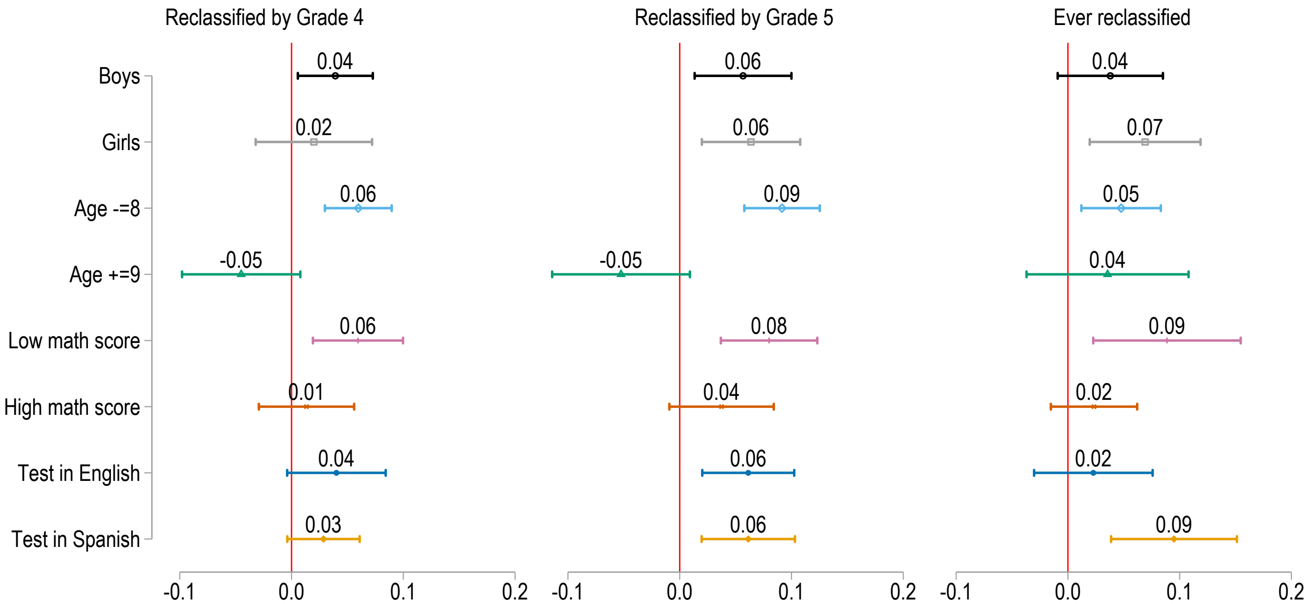
Note: Figure 6 presents non-parametric estimations of the effect of narrowly failing the reading test and accelerated instruction on each outcome. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 7: Effects of Narrowly Failing the Reading Test (ITT) on Academic Achievement by Student Characteristics



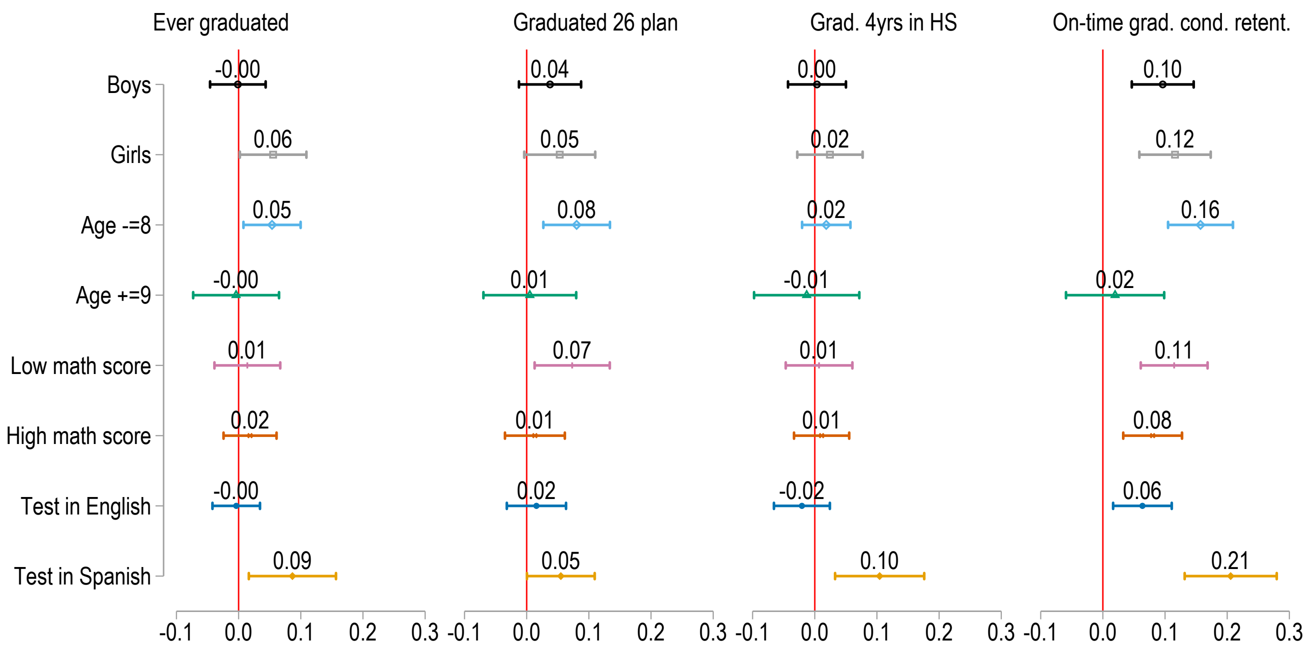
Note: Figure 7 presents point estimates of non-parametric estimations across different student subgroups.The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 8: Effects of Narrowly Failing the Reading Test (ITT) on Reclassification by Student Characteristics



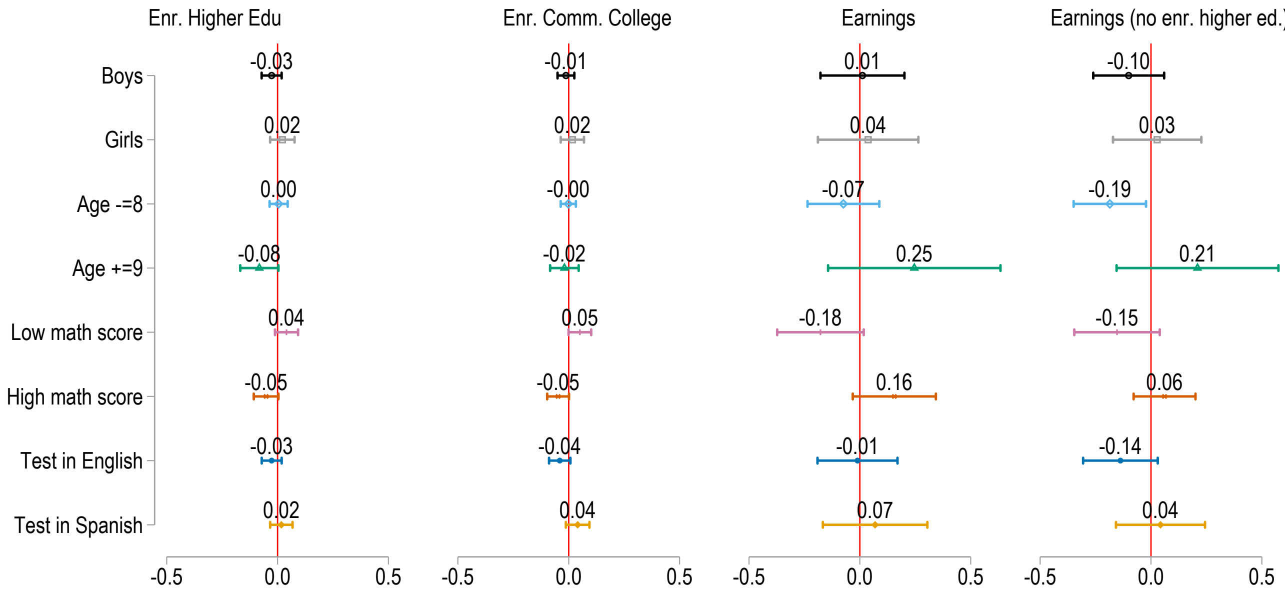
Note: Figure 8 presents point estimates of non-parametric estimations across different student subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 9: Effects of Narrowly Failing the Reading Test (ITT) on High School Completion by Student Characteristics



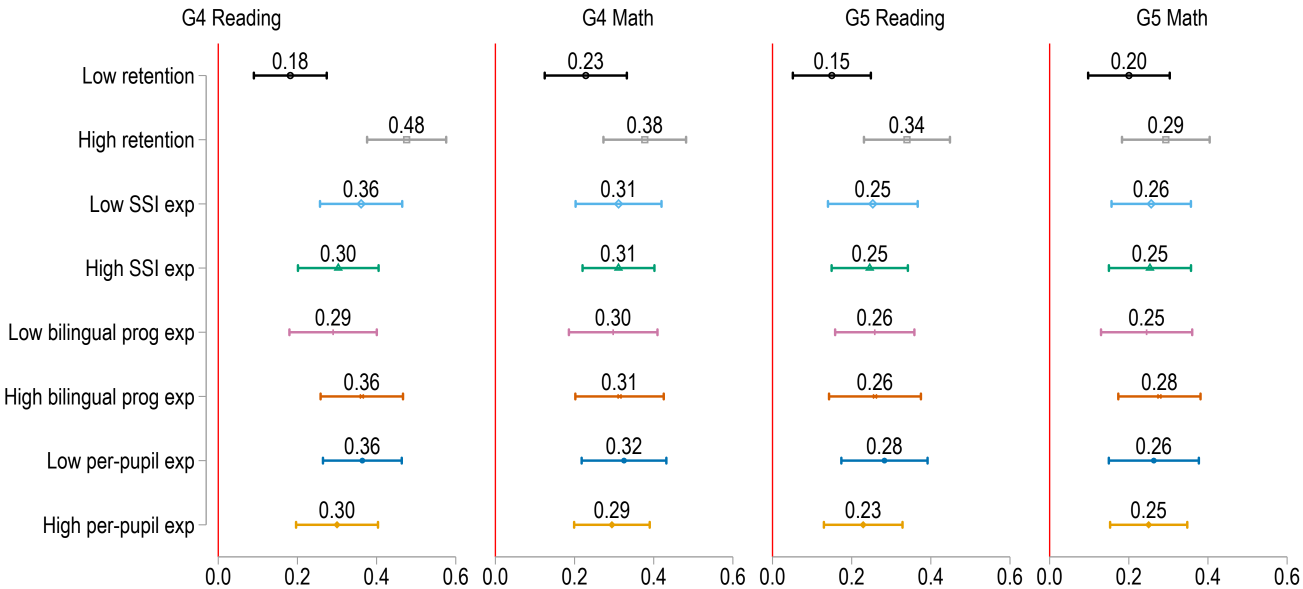
Note: Figure 9 presents point estimates of non-parametric estimations across different student subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 10: Effects of Narrowly Failing the Reading Test (ITT) on Long-Term Outcomes by Student Characteristics



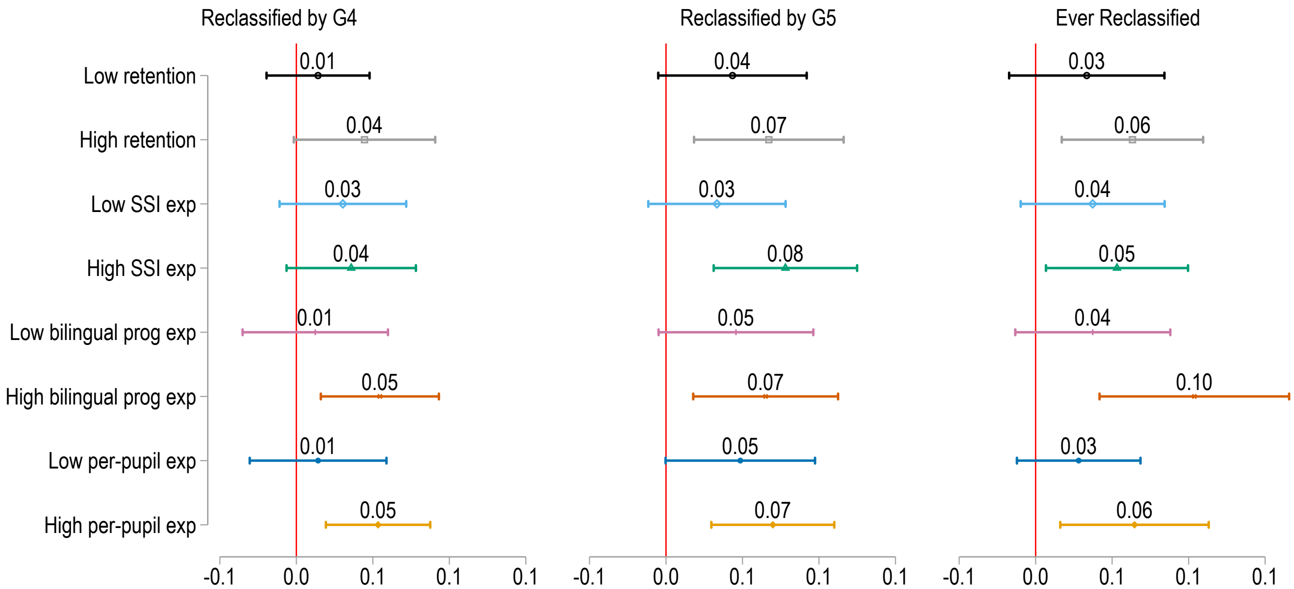
Note: Figure 10 presents point estimates of non-parametric estimations across different student subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 11: Effects of Narrowly Failing the Reading Test (ITT) on Academic Achievement by School Characteristics



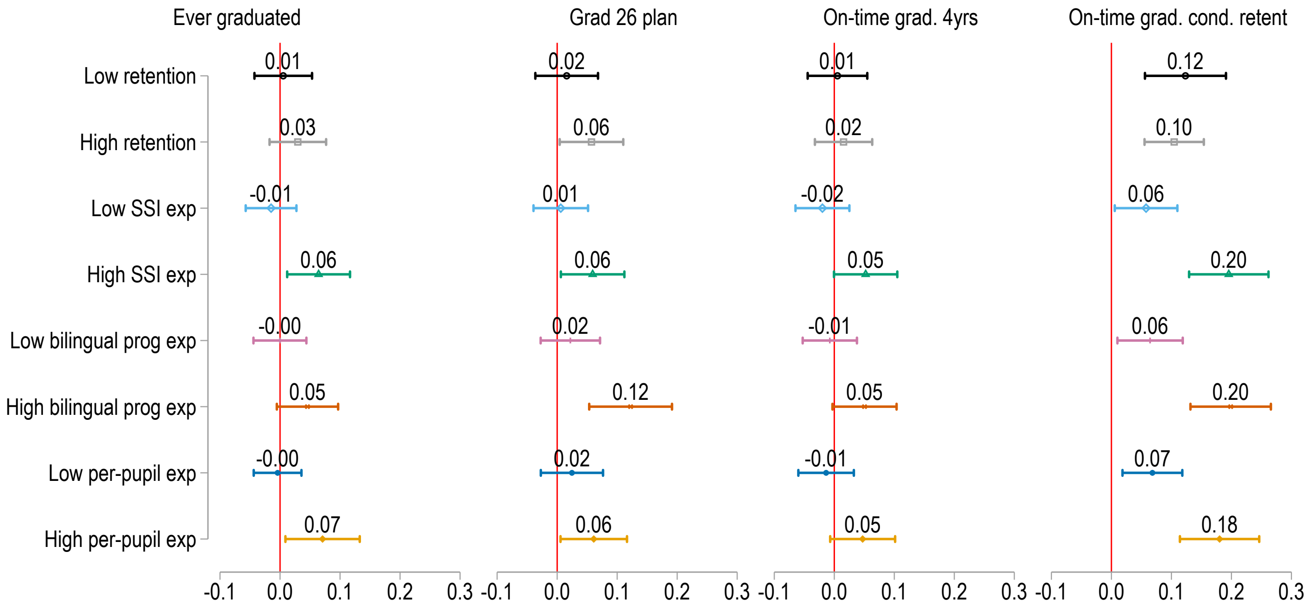
Note: Figure 11 presents point estimates of non-parametric estimations across different school subgroups for each outcome. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 12: Effects of Narrowly Failing the Reading Test (ITT) on Reclassification by School Characteristics



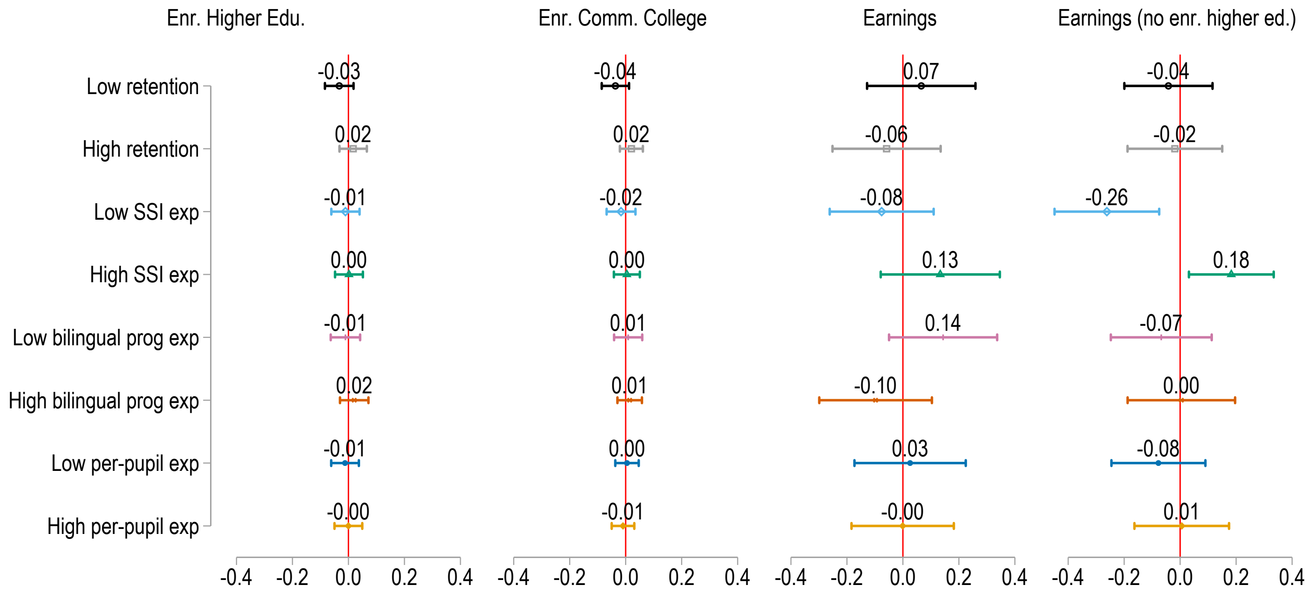
Note: Figure 12 presents point estimates of non-parametric estimations across different school subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 13: Effects of Narrowly Failing the Reading Test (ITT) on High School Completion by School Characteristics



Note: Figure 13 presents point estimates of non-parametric estimations across different school subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure 14: Effects of Narrowly Failing the Reading Test (ITT) on Long-term Outcomes by School Characteristics



Note: Figure 14 presents point estimates of non-parametric estimations across different school subgroups. The estimation includes controls for student and school characteristics and cohort fixed effects. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Table 1: Descriptive Statistics of Third-Grade Students

	Non-MLs	ML-ELs
Panel I: Demographic characteristics		
Male	0.51	0.52
	(0.50)	(0.50)
Econ. disadvantaged	0.49	0.89
	(0.50)	(0.31)
Age	8.14	8.19
	(0.39)	(0.45)
Special education	0.11	0.08
	(0.31)	(0.27)
Asian	0.03	0.05
	(0.17)	(0.21)
Black	0.18	$<$0.01
	(0.38)	(0.09)
Hispanic	0.32	0.93
	(0.47)	(0.25)
Native American	$<$0.01	$<$0.01
	(0.06)	(0.03)
White	0.47	$<$0.01
	(0.50)	(0.11)
N	1,771,241	549,826
Panel II: Academic achievement and grade retention		
Met math standard	0.86	0.76
	(0.35)	(0.42)
N	1,647,263	511,695
Met reading standard (1st attempt)	0.86	0.75
	(0.34)	(0.43)
N	1,723,124	534,801
Retained in Grade 1	0.05	0.07
	(0.22)	(0.25)
Retained in Grade 2	0.03	0.05
	(0.16)	(0.21)
Retained in Grade 3	0.02	0.04
	(0.14)	(0.21)
N	1,771,241	549,826

Note: Table 1 presents descriptive statistics for the full sample of non-EL and ML-EL students who were first-time third-graders from 2002–2003 through 2008–2009. SD are in parentheses. The number of observations for test scores is smaller than for demographic characteristics because not all students have valid scores.

Table 2: School Characteristics of Third-Grade Students

	Non-MLs	ML-ELs
School Enrollment	614.66	683.79
	(215.63)	(209.72)
Panel I: Demographic composition		
Econ. disadvantaged students	0.53	0.78
	(0.28)	(0.21)
Asian students	0.04	0.03
	(0.06)	(0.05)
Black students	0.15	0.11
	(0.19)	(0.14)
Hispanic students	0.40	0.72
	(0.30)	(0.27)
White students	0.42	0.15
	(0.30)	(0.20)
ML-ELs	0.16	0.44
	(0.17)	(0.22)
Bilingual program	0.10	0.34
	(0.16)	(0.23)
N	1,771,241	549,826
Panel II: Academic achievement and grade retention		
Met math standard	0.86	0.80
	(0.11)	(0.12)
N	1,770,381	549,780
Met reading standard (1st attempt)	0.90	0.84
	(0.08)	(0.09)
N	1,770,409	549,784
ML-EL Retention	0.04	0.05
	(0.09)	(0.06)
N	1,609,094	549,662
Non-EL Retention	0.02	0.03
	(0.03)	(0.04)
N	1,771,163	549,419
Panel III: Per-pupil expenditures		
Total	6,890.18	6,710.78
	(223387.89)	(8109.79)

Continued on next page

	Non-MLs	ML-ELs
N	1,770,215	549,734
Bilingual Education	7,610.86	3,624.12
	(33688.30)	(15501.86)
N	800,828	478,622
Accelerated instruction and SSI	238.43	187.29
	(527.22)	(399.44)
N	1,741,869	545,434

Note: Table 2 presents school characteristics for the full sample of non-EL and ML-EL students who were first-time third-graders from 2002–2003 through 2008–2009. SD are in parentheses.

Table 3: Descriptive Statistics of Third-Grade ML-ELs across Test Administrations

	1st attempt	2nd attempt	3rd attempt
Panel I: Demographic characteristics			
Male	0.50	0.56	0.57
	(0.50)	(0.50)	(0.49)
Econ. disadvantaged	0.89	0.93	0.94
	(0.31)	(0.25)	(0.24)
Age	8.16	8.24	8.26
	(0.41)	(0.47)	(0.48)
Asian	0.04	0.02	0.02
	(0.21)	(0.15)	(0.14)
Hispanic	0.94	0.96	0.97
	(0.24)	(0.19)	(0.17)
Home language: Spanish	0.92	0.95	0.96
	(0.26)	(0.22)	(0.20)
Years in EL program in TX	2.89	2.99	3.03
	(0.63)	(0.63)	(0.61)
Newcomer (1-2 years)	0.14	0.12	0.10
	(0.35)	(0.32)	(0.30)
Bilingual program	0.68	0.70	0.70
	(0.47)	(0.46)	(0.46)
ESL program	0.25	0.25	0.24
	(0.43)	(0.43)	(0.43)
Parent denied services	0.12	0.10	0.10
	(0.32)	(0.30)	(0.30)
	475,126	83,178	39,399
Panel II: Acadamic achievement and grade retention			
Met math standard	0.77	0.42	0.31
	(0.42)	(0.49)	(0.46)
	471,277	82,889	39,292
Met reading standard	0.79	0.44	0.39
	(0.40)	(0.50)	(0.49)
	475,126	83,178	39,399
Math raw score	29.97	23.59	21.97
	(7.44)	(7.04)	(6.72)
	471,599	83,176	39,399

Continued on next page

	1st attempt	2nd attempt	3rd attempt
Reading raw scores	26.97	20.76	20.14
	(7.71)	(6.19)	(6.02)
	475,126	83,178	39,399
Took reading test in Spanish	0.37	0.38	0.37
	(0.48)	(0.49)	(0.48)
	475,126	83,178	39,399
Retained in Grade 1	0.06	0.10	0.11
	(0.23)	(0.30)	(0.32)
Retained in Grade 2	0.04	0.07	0.07
	(0.21)	(0.25)	(0.26)
Retained in Grade 3	0.05	0.22	0.34
	(0.21)	(0.42)	(0.47)
	475,126	83,178	39,399
Panel III:English profiency			
Reclassified	0.27	0.10	0.05
	(0.44)	(0.30)	(0.21)
	475,126	83,178	39,399
Advanced English language profiency			
Writing	0.17	0.02	<0.01
	(0.37)	(0.15)	(0.12)
Speaking	0.26	0.09	0.07
	(0.44)	(0.28)	(0.26)
Listening	0.33	0.12	0.10
	(0.47)	(0.32)	(0.30)
Reading	0.48	0.07	0.02
	(0.50)	(0.26)	(0.16)
All domains	0.13	0.02	<0.01
	(0.34)	(0.12)	(0.10)
N	347,842	62,085	30,848

Note: Table 3 presents descriptive statistics of ML-ELs across test administrations. SD are in parentheses. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. Students who took the third administration failed the first and second administrations of TAKS. TELPAS scores are available from 2005.

Table 4: Effect of Third-Grade Retention on Student Academic Achievement

	(1) ITT	(2) TOT
Panel I: Grade 4		
Reading	0.333***	0.939***
	(0.037)	(0.078)
Mean	-1.12	-1.11
Optimal BW	3.81	5.22
N	14520	21411
Math	0.309***	0.933***
	(0.039)	(0.083)
Mean	-.89	-.88
Optimal BW	3.64	5.44
N	14650	21614
Writing	0.175***	0.505***
	(0.025)	(0.054)
Mean	-.87	-.87
Optimal BW	3.94	5.95
N	14621	21570
Panel II: Grade 5		
Reading	0.250***	0.758***
	(0.040)	(0.085)
Mean	-1.08	-1.06
Optimal BW	3.77	5.89
N	13342	19538
Math	0.258***	0.778***
	(0.037)	(0.087)
Mean	-.84	-.82
Optimal BW	4.24	6.14
N	16698	22380

Note: Table 4 presents point estimates of non-parametric estimation of the effect of retention and accelerated instruction on each outcome. Reading and math raw scores are standardized within year, subject, and grade. The estimation includes controls for student and school characteristics and cohort fixed effects. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and whose scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parentheses, are clustered at the school level.*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 5: Effect of Third-Grade Retention on Reclassification

	(1) ITT	(2) TOT
Years classified as EL	-0.251***	-0.636***
	(0.095)	(0.239)
Mean	5.88	5.78
Optimal BW	4.486	6.077
N	19610	26507
Ever reclassified	0.049***	0.109***
	(0.017)	(0.039)
Mean	.72	.73
Optimal BW	3.405	5.281
N	15701	23240
Panel II: Same-grade comparisons		
Reclassified by Grade 4	0.033**	0.129***
	(0.015)	(0.032)
Mean	.2	.21
Optimal BW	3.407	6.081
N	15701	26507
Reclassified by Grade 5	0.059***	0.181***
	(0.016)	(0.039)
Mean	.31	.32
Optimal BW	4.232	5.386
N	19610	23240
Panel III: Same-age comparisons		
Reclassified 1 year after retention	-0.012	-0.033
	(0.012)	(0.030)
Mean	.18	.18
Optimal BW	4.437	5.99
N	19610	23240
Reclassified 2 years after retention	0.032**	0.096***
	(0.015)	(0.037)
Mean	.29	.29
Optimal BW	4.304	5.683
N	19212	22750

Note: Table 5 presents point estimates of non-parametric estimation of the effect of retention and accelerated instruction on each outcome. The estimation includes controls for student and school characteristics and cohort fixed effects. Years classified as EL is the number of years a student receive services as a EL since their first time in third grade. In Panel II, reclassification is measured as whether the student has been reclassified by Grade 4 or Grade 5. In Panel III, reclassification is measured as whether a student attained reclassification 1 or 2 years after their first time in third grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and whose scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002–2003 through 2008–2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table 6: Effect of Third-Grade Retention on High School Completion

	(1) ITT	(2) TOT
Ever Graduated	0.019	0.022
	(0.018)	(0.038)
Mean	.74	.75
Optimal BW	3.36	5.59
N	14746	21800
Graduated with 26 credit plan	0.045**	0.056
	(0.020)	(0.043)
Mean	.571	.58
Optimal BW	3.24	5.54
N	14746	21800
On-time graduation (4 years in high school)	0.008	0.024
	(0.017)	(0.042)
Mean	.69	.71
Optimal BW	3.69	5.25
N	14290	21098
On-time graduation (Conditional on retention)	0.122***	0.231***
	(0.023)	(0.040)
Mean	.55	.57
Optimal BW	3.03	6.09
N	14746	24889
Dropout	-0.007	-0.021
	(0.013)	(0.030)
Mean	.13	.12
Optimal BW	3.62	5.43
N	14746	21800

Note: Table 6 presents point estimates of non-parametric estimation of the effect of retention and accelerated instruction on each outcome. The estimation includes controls for student and school characteristics and cohort fixed effects. The Optimal BW rows show the bandwidth used for each estimation. High school graduation is defined as whether a student has received a standard high school degree by the end of 2022. On-time graduation conditional on retention is a binary indicator for whether the student graduated by the spring of the 9th year after being promoted from third grade. Students who were still enrolled in school after 9 years or who dropped out of high school were coded as "nongraduates". Dropout is defined as whether a student in the administrative data is formally identified as a dropout by the end of the 2022 year and does not have a graduation date. Students identified as dropouts but re-enrolling in schools and having a graduation date are considered high school completers. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and whose scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. Standard errors, shown in parenthesis, are clustered at the school level.*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 7: Effect of Third-Grade Retention on Post-Secondary Enrollment

	(1) ITT	(2) TOT
Panel I: Enrollment at ages 17 to 20		
Higher education enrollment	-0.002	0.006
	(0.016)	(0.041)
Mean	.37	.37
Optimal BW	4.347	5.444
N	17999	21298
Community college enrollment	-0.002	0.004
	(0.015)	(0.039)
Mean	.31	.31
Optimal BW	4.392	5.765
N	17999	21298
Panel II: Enrollment 9-12 years post 3rd grade		
Higher education enrollment	-0.004	-0.002
	(0.016)	(0.043)
Mean	.37	.38
Optimal BW	4.395	5.351
N	17663	20892
Community college enrollment	-0.007	-0.017
	(0.015)	(0.040)
Mean	.31	.32
Optimal BW	4.512	5.561
N	17663	20892

Note: Table 7 presents point estimates of non-parametric estimation of the effect of retention and accelerated instruction on each outcome. The estimation includes controls for student and school characteristics and cohort fixed effects. Post-secondary enrollment at any institution includes enrollment in 2-year public institutions, 4-year public institutions, private institutions, and for-profit institutions in Texas. Panel I measures enrollment by age, whether the student was enrolled in post-secondary institutions at ages 17 to 20. Panel II measures enrollment by years after completing third-grade, whether a student was enrolled nine to 12 years after being promoted from third grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test, and whose scores were within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. Standard errors, shown in parenthesis, are clustered at the school level.*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table 8: Effect of Third-Grade Retention on Earnings at Ages 20-22

	(1) ITT	(2) TOT
Average Quarterly Earnings	0.010	-0.046
	(0.072)	(0.167)
Mean	8.15	8.15
Optimal BW	3.9	5.52
N	11388	16817
Average Quarterly Earnings (not enrolled in higher education)	-0.039	-0.096
	(0.053)	(0.127)
Mean	8.38	8.38
Optimal BW	4.28	6.19
N	10317	13987

Note: Table 8 presents point estimates of non-parametric estimation of the effect of retention and accelerated instruction on average quarterly earnings. Earnings are expressed in 2022 dollars. A logarithmic transformation was applied. The estimation includes controls for student and school characteristics and cohort fixed effects. The first row shows the effects of third-grade retention on average earnings for all ML-ELs in the sample at ages 20 to 22. The second row show the effects of grade retention for ML-ELs in the sample who were not enrolled in post-secondary institutions at ages 20 to 22. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and whose scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

References

Abedi, J. (2006). Psychometric Issues in the ELL Assessment and Special Education Eligibility - Jamal Abedi, 2006. *Teachers College Record*, 108(11):2282–2303.

Abraham, K. G., Haltiwanger, J. C., Sandusky, K., and Spletzer, J. R. (2017). Measuring the Gig Economy: Current Knowledge and Open Issues. In *Measuring and Accounting for Innovation in the Twenty-First Century*, pages 257–298. University of Chicago Press, Chicago, IL.

Anderson, G. E., Jimerson, S. R., and Whipple, A. D. (2005). Student Ratings of Stressful Experiences at Home and School: Loss of a Parent and Grade Retention as Superlative Stressors. *Journal of Applied School Psychology*, 21(1):1–20.

Backes, B., Cowan, J., Goldhaber, D., and Theobald, R. (2024). How to Measure a Teacher: The Influence of Test and Nontest Value-Added on Long-Run Student Outcomes. *Journal of Human Resources*, pages 1023–13180R2.

Beuermann, D. W. and Jackson, C. K. (2020). The Short and Long-Run Effects of Attending The Schools that Parents Prefer. *Journal of Human Resources*. Publisher: University of Wisconsin Press Section: Article.

Brooks, M. D. (2019). *Transforming Literacy Education for Long-Term English Learners: Recognizing Brilliance in the Undervalued*. Routledge, New York.

Callahan, R. M. (2005). Tracking and High School English Learners: Limiting Opportunity to Learn. *American Educational Research Journal*, 42(2):305–328.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation Testing Based on Density Discontinuity. *The Stata Journal: Promoting communications on statistics and Stata*, 18(1):234–261.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star *. *The Quarterly Journal of Economics*, 126(4):1593–1660.

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.

Curry, J. (2004). Aisd k-4 accelerated reading and mathematics instruction evaluation, 2003-2004. Technical report, Austin Independent School District, Office of Accountability.

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. Publisher: Routledge _eprint: https://doi.org/10.1080/10824669.2014.954044.

Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market*. *The Quarterly Journal of Economics*, 132(4):1593–1640.

Deming, D. J., Hastings, J. S., Kane, T. J., and Staiger, D. O. (2014). School Choice, School Quality, and Postsecondary Attainment. *American Economic Review*, 104(3):991–1013.

Dondero, M. and Muller, C. (2012). School Stratification in New and Established Latino Destinations. *Social Forces*, 91(2):477–502.

Eren, O., Depew, B., and Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics*, 153:9–31.

Eren, O., Lovenheim, M. F., and Mocan, H. N. (2022). The Effect of Grade Retention on Adult Crime: Evidence from a Test-Based Promotion Policy. *Journal of Labor Economics*, 40(2):361–395.

Figlio, D. and Özek, U. (2020). An extra year to learn English? Early grade retention and the human capital development of English learners. *Journal of Public Economics*, 186:104184.

Flores, N., Kleyn, T., and 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.

Garcia, E. B., Sulik, M. J., and Obradović, J. (2019). Teachers' perceptions of students' executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology*, 111(5):918–931.

Goldhaber, D. and Özek, U. (2019). How Much Should We Rely on Student Test Achievement as a Measure of Success? *Educational Researcher*, 48(7):479–483. Publisher: American Educational Research Association.

Goos, M., Pipa, J., and Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review*, 34:100401.

Greene, J. P. and Winters, M. A. (2007). Revisiting Grade Retention: An Evaluation of Florida's Test-Based Promotion Policy. *Education Finance and Policy*, 2(4):319–340.

Gàndara, P. and Baca, G. (2008). NCLB and Califonia's English Language Learners: The perfect storm. *Lang Policy*, 7:201–216.

Gándara, P., Rumberger, R., Maxwell-Jolly, J., and Callahan, R. (2003). English Learners in California Schools: Unequal resources, 'Unequal outcomes. *Education Policy Analysis Archives*, 11:36–36.

Gándara, P. and Rumberger, R. W. (2009). Immigration, Language, and Education: How Does Language Policy Structure Opportunity? *Teachers College Record*, 111(3):750–782.

Hotz, J. and Scholz, J. (2002). *Studies of Welfare Populations: Data Collection and Research Issues.* National Research Council: National Academies Press.

Hwang, N. and Koedel, C. (2023). Helping or Hurting: The Effects of Retention in the Third Grade on Student Outcomes. *Educational Evaluation and Policy Analysis*, XX(X):1–24.

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

Jacob, B. A. and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of economics and statistics*, 86(1):226–244.

Jacob, B. A. and Lefgren, L. (2009). The Effect of Grade Retention on High School Completion. *American Economic Journal: Applied Economics*, 1(3):33–58.

Jimerson, S. R. (2001). A Synthesis of Grade Retention Research: Looking Backward and Moving Forward. *The California School Psychologist*, 6(1):47–59.

Katz, S. R. (1999). Teaching in Tensions: Latino Immigrant Youth, Their Teachers, and the Structures of Schooling.

Larsen, M. F. and Valant, J. (2023). The Long-Term Effects of Grade Retention: Evidence on Persistence through High School and College. *Journal of Research on Educational Effectiveness*, pages 1–32.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

Lovenheim, M. F. and Willén, A. (2019). The Long-Run Effects of Teacher Collective Bargaining. *American Economic Journal: Economic Policy*, 11(3):292–324.

Ludwig, J. and Miller, D. L. (2007). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design*. *The Quarterly Journal of Economics*, 122(1):159–208.

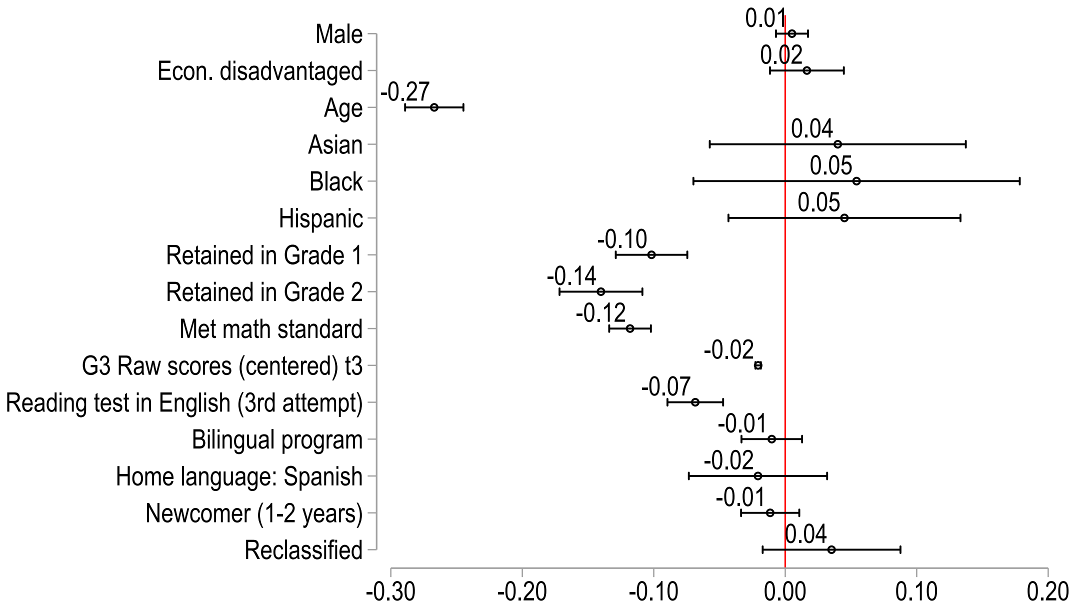
Mariano, L. T. and Martorell, P. (2013). The Academic Effects of Summer Instruction and Retention in New York City. *Educational Evaluation and Policy Analysis*, 35(1):96–117.

Mariano, L. T., Martorell, P., and Berglund, T. (2024). The Effects of Grade Retention on High School Outcomes: Evidence from New York City Schools. *Journal of Research on Educational*

Effectiveness, pages 1–31.
Milner, H. R. (2012). Beyond a Test Score: Explaining Opportunity Gaps in Educational Practice. *Journal of Black Studies*, 43(6):693–718.
Mumma, K. S. and Winters, M. A. (2023). The Effect of Retention Under Mississippi's Test-Based Promotion Policy.
Murphy, A. F. and Torff, B. (2019). Teachers' Beliefs About Rigor of Curriculum for English Language Learners. *The Educational Forum*, 83(1):90–101. Publisher: Routledge _eprint: https://doi.org/10.1080/00131725.2018.1505991.
NCES (2023). English Learners in Public Schools.
Ochoa, G. L. (2013). *Academic Profiling: Latinos, Asian Americans, and the Achievement Gap*. U of Minnesota Press.
Olsen, L. (1997). *Made in America: Immigrant Students in Our Public Schools*. The New Press, 450 West 41st Street, New York, NY 10036 ($14. ERIC Number: ED421569.
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.
Quintero, D. and Hansen, M. (2021). As we tackle school segregation, don't forget about English Learner students.
Redding, C. and Carlo, S. M. (2023). The Adoption of Test-Based Grade Retention Policies: An Event History Analysis. *Educational Policy*, 38(4).
Robinson-Cimpian, J. P., Thompson, K. D., and Umansky, I. M. (2016). Research and Policy Considerations for English Learner Equity. *Policy Insights from the Behavioral and Brain Sciences*, 3(1):129–137. Publisher: SAGE Publications.
Ruiz, R. (1984). Orientations in Language Planning. *ilingual Research Journal*, 8(2):15–34.
Schwerdt, G., West, M. R., and Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, 152:154–169.
Shepard, L. and Smith, M. (2017). Synthesis of research on grade retention. *Educational Leadership*, 47(8):84–88.
Soland, J. and Sandilos, L. E. (2021). English Language Learners, Self-efficacy, and the Achievement Gap: Understanding the Relationship between Academic and Social-Emotional Growth. *Journal of Education for Students Placed at Risk (JESPAR)*, 26(1):20–44.
TEC §28.0211 (1999). Texas school law bulletin.
Umansky, I. M. (2016). Leveled and Exclusionary Tracking: English Learners' Access to Academic Content in Middle School. *American Educational Research Journal*, 53(6):1792–1833.
U.S. Department of Education (2016). Educational Experiences of English Learners: Grade Retention, High School Graduation, and GED Attainment, 2011-12.
Valdes, G. (2004). *Learning and Not Learning English: Latino Students in American Schools*. Teachers College Press. Google-Books-ID: wTGRDQAAQBAJ.
Weiss, A. (1995). Human Capital vs. Signalling Explanations of Wages. *Journal of Economic Perspectives*, 9(4):133–154.
Özek, U. and Mariano, L. (2023). Can grade retention help with covid-19 learning recovery in schools?. Technical report, Brookings Institution.

A Appendix

Figure A.1: Probability of Retention Conditional on Failing the Reading Test (Compliers)

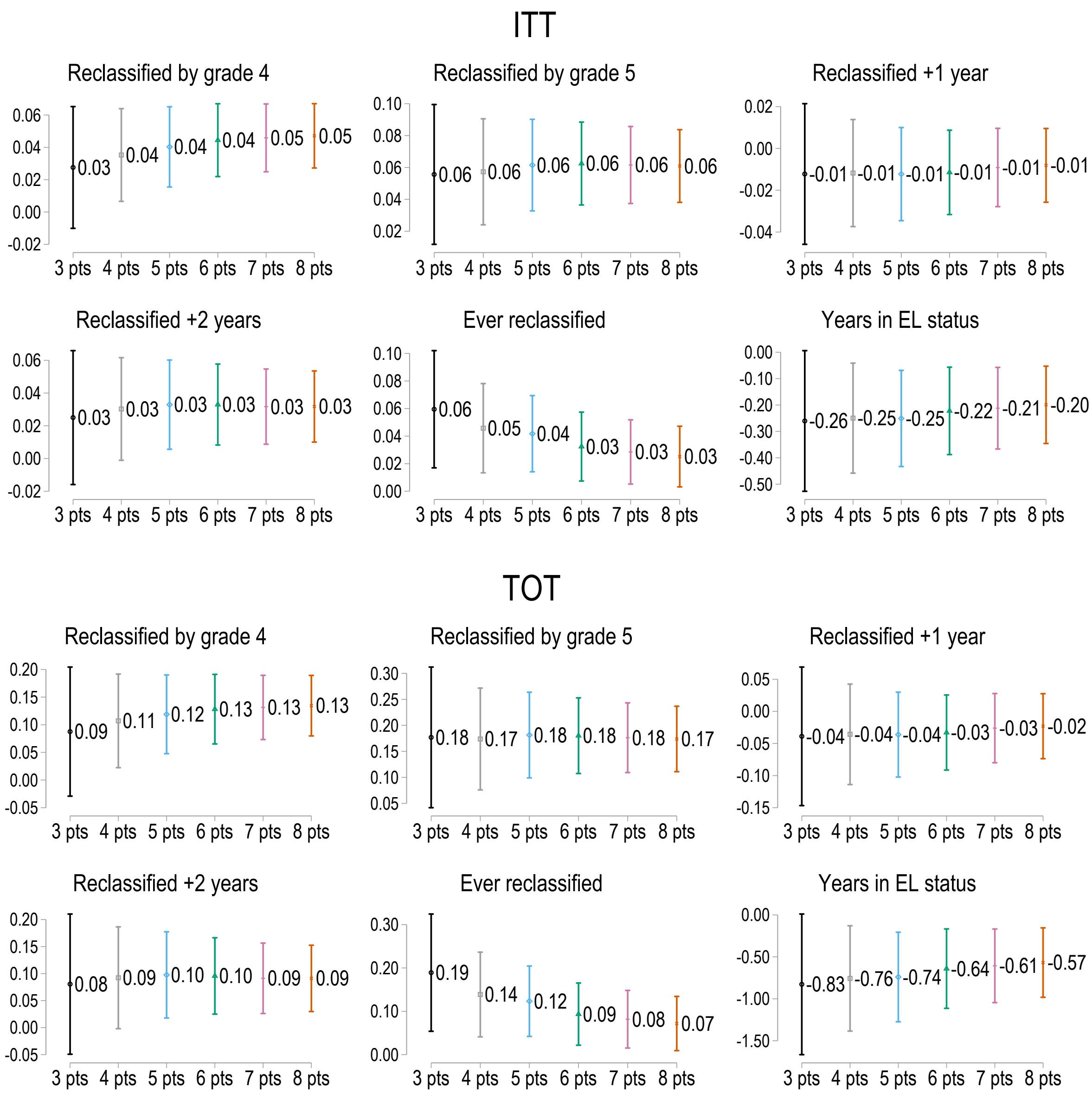


Note: Figure A.1 presents point estimates of an OLS regression where the dependent variable is being a complier (a student who failed the third attempt of the reading test and was retained). The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.2: Effects of Third-Grade Retention on Academic Achievement across Several Bandwidths

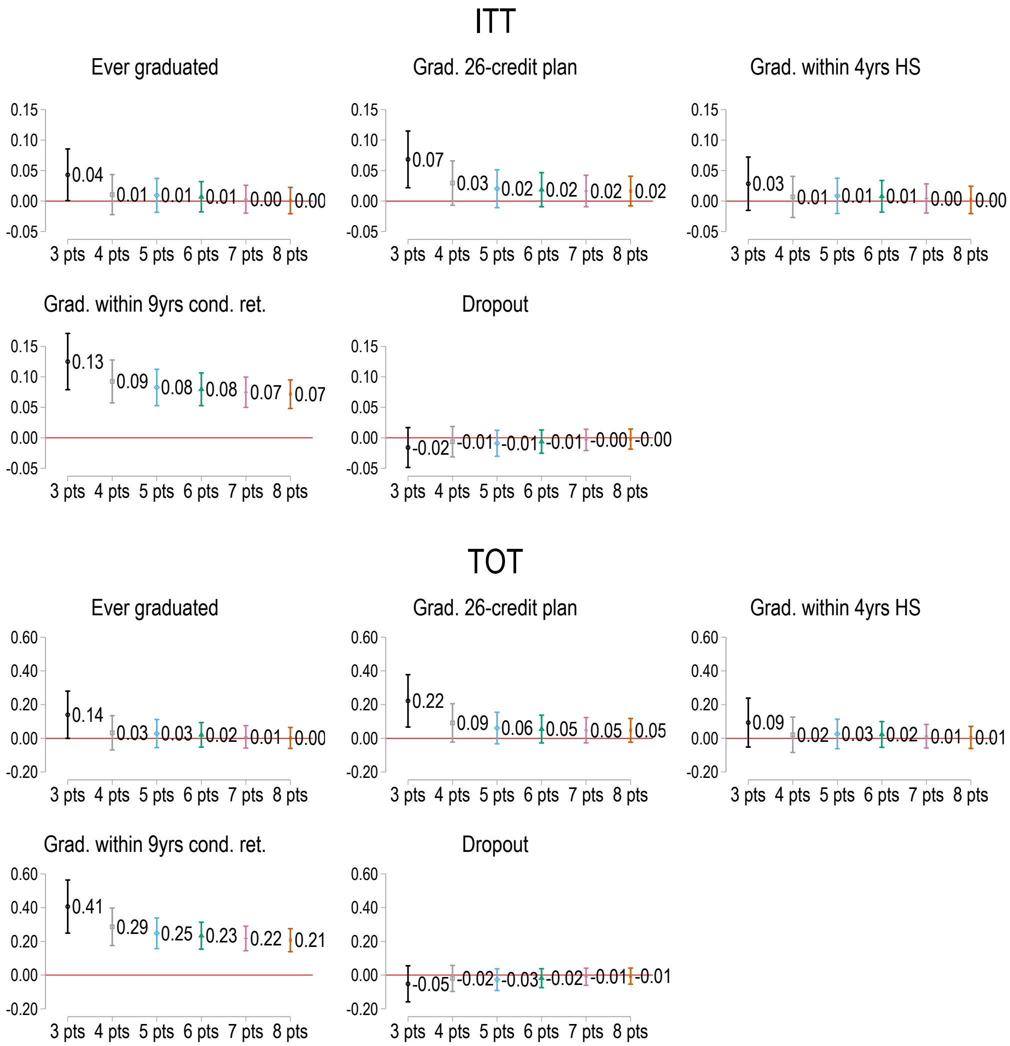
Note: Figure A.2 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome across different bandwidths. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.3: Effects of Third-Grade Retention on Reclassification across Several Bandwidths



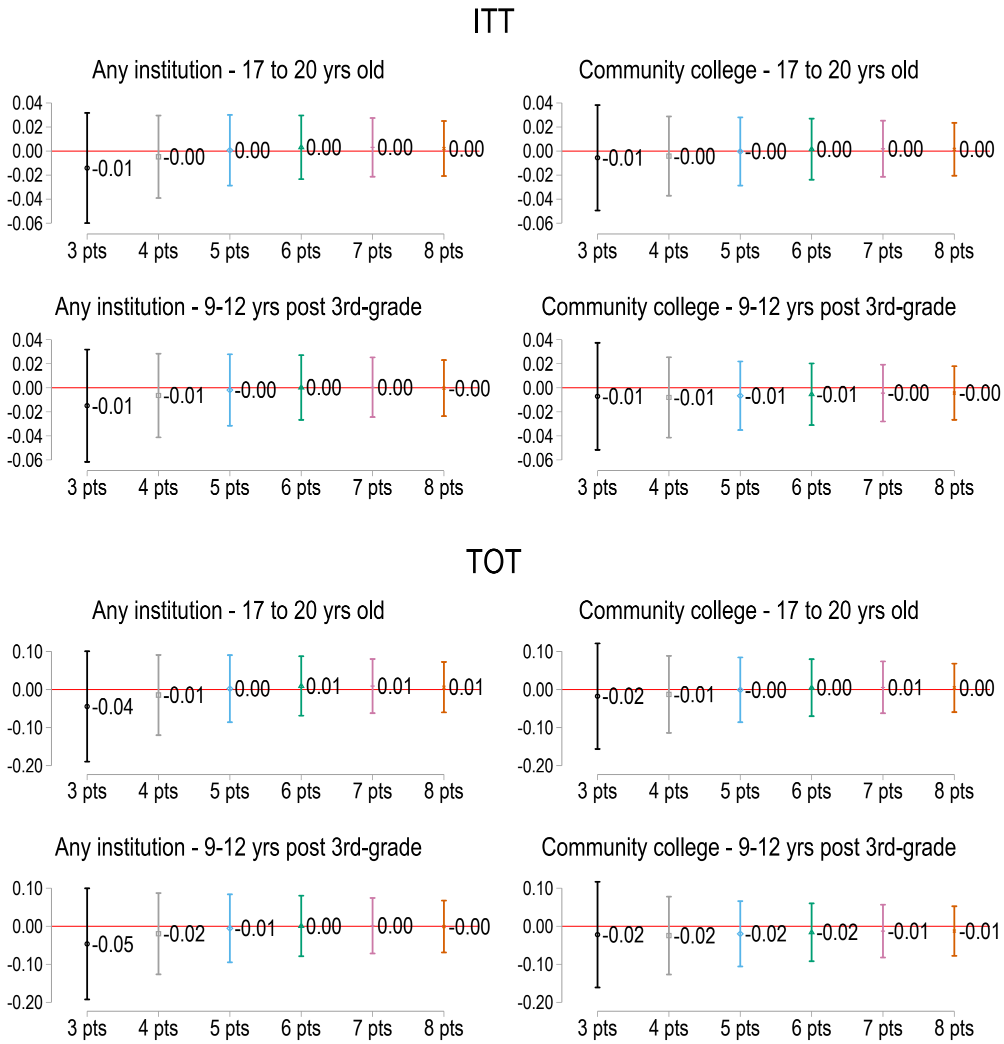
Note: Figure A.3 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome across different bandwidths. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.4: Effects of Third-Grade Retention on High School Completion across Several Bandwidths



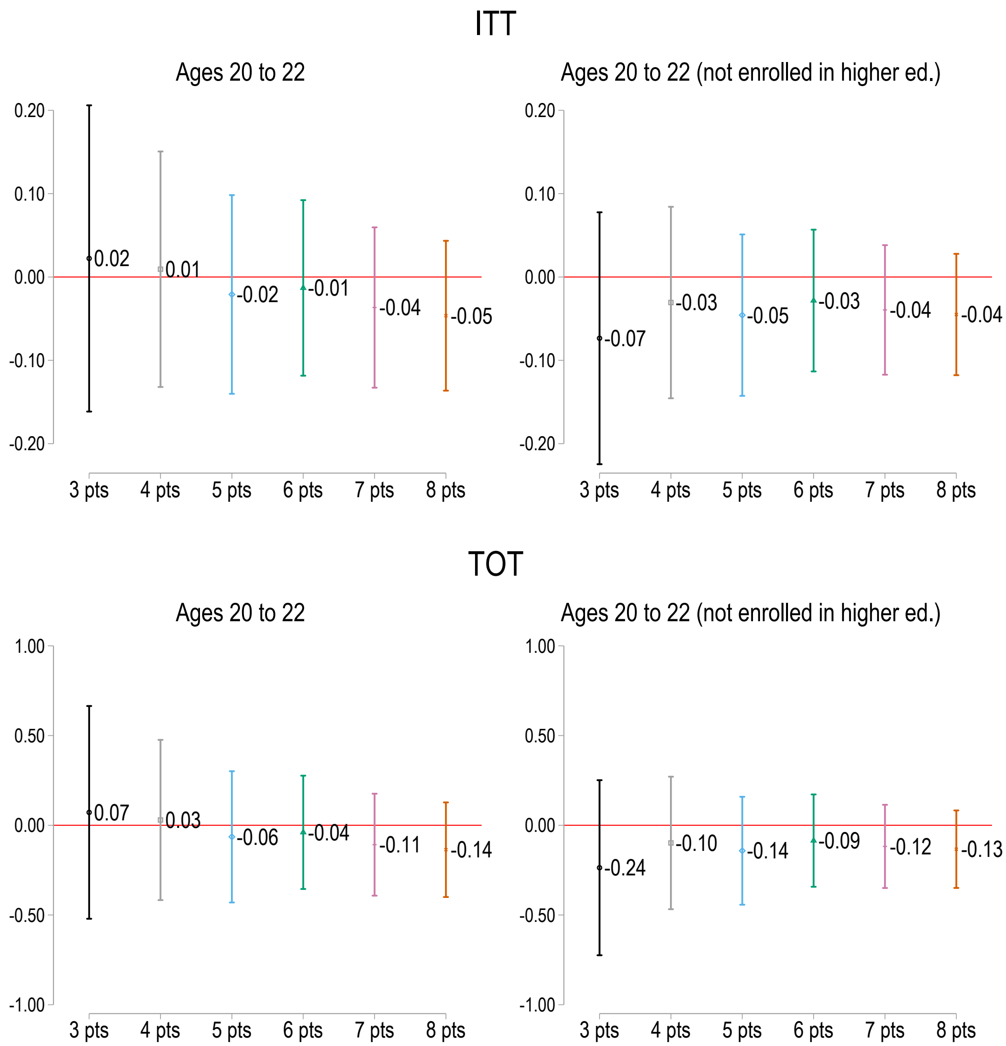
Note: Figure A.4 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome across different bandwidths. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.5: Effects of Third-Grade Retention on Post-Secondary Enrollment across Several Bandwidths



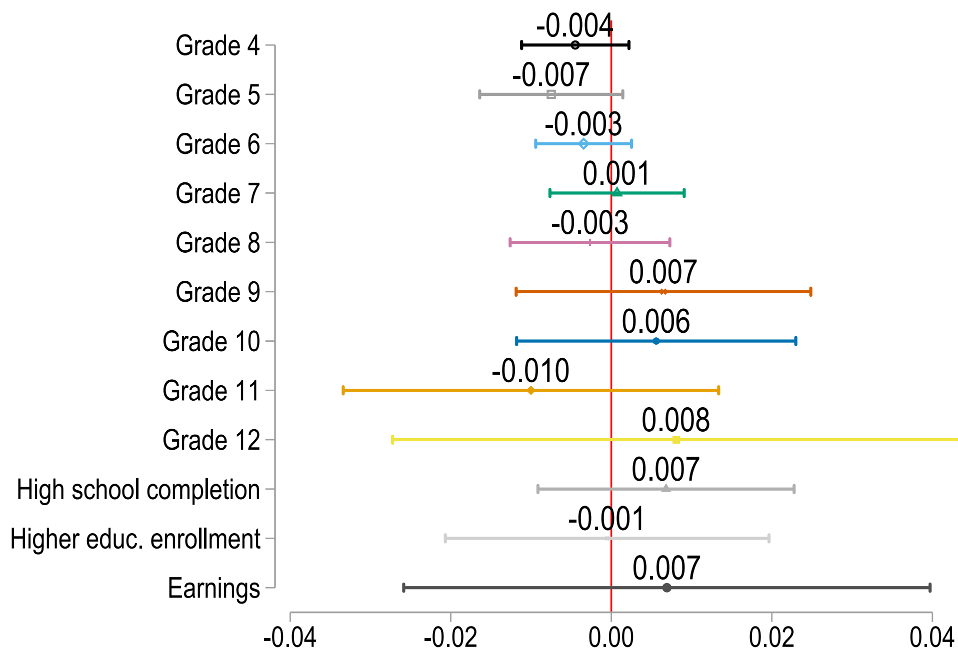
Note:Figure A.5 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome across different bandwidths. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.6: Effects of Third-Grade Retention on Earnings across Several Bandwidths



Note: Figure A.6 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome across different bandwidths. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Figure A.7: Probability of Differential Attrition



Note: Figure A.7 presents point estimates of non-parametric estimations. The dependent variable is whether the student has non-missing information for each grade or outcome. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment.

Table A.1: Effect of Third-Grade Retention on Student Academic Achievement (No Controls)

	(1) ITT	(2) TOT
Panel I: Grade 4		
Reading	0.344***	0.940***
	(0.038)	(0.077)
Mean	-1.12	-1.11
Optimal BW	3.7	5.6
N	14549	21453
Math	0.335***	0.986***
	(0.043)	(0.102)
Mean	-.89	-.88
Optimal BW	3.85	5.08
N	14673	21647
Writing	0.183***	0.514***
	(0.023)	(0.055)
Mean	-.86	-.86
Optimal BW	4.78	6.16
N	18273	24569
Panel II: Grade 5		
Reading	0.259***	0.772***
	(0.041)	(0.088)
Mean	-1.08	-1.04
Optimal BW	3.82	6.03
N	13367	22231
Math	0.282***	0.823***
	(0.046)	(0.096)
Mean	-.85	-.831
Optimal BW	3.83	6.48
N	13460	22418

Note: Table A.1 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. The estimation includes only controls for cohort. Reading and math raw scores are standarized within year, subject, and grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A.2: Effect of Third-Grade Retention on Reclassification (No Controls)

	(1) ITT	(2) TOT
Years classified as EL	-0.265***	-0.671**
	(0.103)	(0.265)
Mean	5.88	5.82
Optimal BW	4.486	5.819
N	19661	23302
Ever reclassified	0.053***	0.121***
	(0.018)	(0.042)
Mean	.72	.73
Optimal BW	3.393	5.162
N	15743	23302
Panel II: Same-grade comparisons		
Reclassified by Grade 4	0.037**	0.132***
	(0.016)	(0.033)
Mean	.2	.21
Optimal BW	3.618	6.372
N	15743	26575
Reclassified by Grade 5	0.061***	0.186***
	(0.018)	(0.042)
Mean	.31	.32
Optimal BW	3.852	5.344
N	15743	23302
Panel III: Same-age comparisons		
Reclassified by 1 year after retention	-0.011	-0.032
	(0.013)	(0.031)
Mean	.18	.19
Optimal BW	4.057	6.056
N	19661	26575
Reclassified by 2 years after retention	0.032*	0.099**
	(0.017)	(0.040)
Mean	.28	.29
Optimal BW	3.909	5.546
N	15426	22810

Note: Table A.2 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. The estimation includes only controls for cohort. Years classified as EL is the number of years a student received services as a EL since their first time in third grade. In Panel II, reclassification is measured as whether the student has been reclassified by grade 4 or grade 5. In Panel III, reclassification is measured as whether a student attained reclassification 1 or 2 years after their first time in third grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth.The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table A.3: Effect of Third-Grade Retention on High School Completion (No Controls)

	(1) ITT	(2) TOT
Ever Graduated	0.027	0.022
	(0.019)	(0.040)
Mean	.74	.75
Optimal BW	3.24	5.6
N	14786	21859
Graduated with a 26 credit plan	0.061***	0.057
	(0.022)	(0.045)
Mean	.571	.58
Optimal BW	3.11	5.61
N	14786	21859
On-time graduation (within 4 years in high school)	0.013	0.028
	(0.018)	(0.043)
Mean	.69	.7
Optimal BW	3.65	5.23
N	14330	21157
On-time graduation (conditional on retention)r	0.132***	0.241***
	(0.025)	(0.046)
Mean	.55	.56
Optimal BW	2.92	5.17
N	10801	21859
Dropout	-0.010	-0.020
	(0.013)	(0.030)
Mean	.13	.12
Optimal BW	3.56	5.62
N	14786	21859

Note: Table A.3 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. The estimation includes only controls for cohort. The Optimal BW rows show the bandwidth used for each estimation. High school graduation is defined as whether a student has received a standard high school degree by the end of 2022. On-time graduation is as a binary indicator for whether the student graduated by the spring of the 9th year after being promoted from third grade. Students who were still enrolled in school after 9 years or who dropped out of high school were coded as "nongraduates". Dropout is defined as whether a student in the administrative data is formally identified as a dropout by the end of the 2022 year and does not have a graduation date. Students identified as dropouts but re-enrolling in schools and having a graduation date are considered high school completers. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table A.4: Effect of Third-Grade Retention on Post-secondary Enrollment (No Controls)

	(1) ITT	(2) TOT
Panel I: Enrollment at ages 17 to 20		
Enrollment in higher education	0.001	0.009
	(0.016)	(0.044)
Mean	.37	.37
Optimal BW	4.48	5.454
N	18044	21353
Community colleges	0.001	0.007
	(0.015)	(0.040)
Mean	.31	.31
Optimal BW	4.561	5.78
N	18044	21353
Panel II: Enrollment 9-12 years post 3rd-grade completion		
Enrollment in higher education	-0.003	-0.002
	(0.016)	(0.045)
Mean	.37	.37
Optimal BW	4.47	5.422
N	17706	20945
Community colleges	-0.006	-0.016
	(0.015)	(0.041)
Mean	.31	.32
Optimal BW	4.57	5.632
N	17706	20945

Note: Table A.4 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. The estimation includes only controls for cohort. Post-secondary enrollment at any institution includes enrollment in 2-year public institutions, 4-year public institutions, private institutions, and for-profit institutions in Texas. Panel I measures enrollment by age, whether the student was enrolled in post-secondary institutions at ages 17 to 20. Panel II measures enrollment by years after completing third-grade, whether a student was enrolled 10 to 12 years after being promoted from third grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table A.5: Effect of Third-Grade Retention on Earnings at Ages 20-22 (No Controls)

	(1) ITT	(2) TOT
Average Quarterly Earnings	0.001	-0.049
	(0.075)	(0.171)
Mean	8.14	8.15
Optimal BW	3.9	5.59
N	11418	16861
Average Quarterly Earnings(cond. not enrolled in higher education)	-0.038	-0.079
	(0.055)	(0.139)
Mean	8.381	8.3
Optimal BW	4.33	5.91
N	10343	12285

Note: Table A.5 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. The estimation includes only controls for cohort. Earnings are expressed in 2022 dollars. A logarithmic transformation was applied. The first row shows the effects of third-grade retention on average earnings for all ML-ELs in the sample at ages 20 to 22. The second row show the effects on grade retention for ML-ELs in the sample who were not enrolled in post-secondary institutions at ages 20 to 22. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table A.6: Effect of Third-grade Retention on Reclassification (Grades 6 to 11)

	(1) ITT	(2) ITT	(3) TOT	(4) TOT
Grade 6	0.062***	0.059***	0.174***	0.167***
	(0.018)	(0.017)	(0.044)	(0.041)
Mean	.42	.42	.43	.43
Optimal BW	4.1	4.07	5.36	5.36
N	19661	19610	23240	23240
Grade 7	0.049**	0.046**	0.125***	0.118***
	(0.019)	(0.018)	(0.044)	(0.041)
Mean	.51	.51	.53	.53
N	15743	15701	23240	23240
Grade 8	0.056***	0.050***	0.111**	0.105***
	(0.020)	(0.018)	(0.044)	(0.040)
Mean	.61	.61	.62	.62
Optimal BW	3.31	3.45	5.62	5.62
N	15743	15701	23240	23240
Grade 9	0.056***	0.051***	0.125***	0.118***
	(0.019)	(0.018)	(0.043)	(0.041)
Mean	.65	.65	.66	.66
Optimal BW	3.88	3.97	5.5	5.5
N	15743	15701	23240	23240
Grade 10	0.057***	0.052***	0.132***	0.125***
	(0.018)	(0.017)	(0.043)	(0.040)
Mean	.69	.69	.70	.70
Optimal BW	3.78	3.86	5.37	5.37
N	15743	15701	23240	23240
Grade 11	0.052***	0.048***	0.119***	0.109***
	(0.018)	(0.017)	(0.042)	(0.039)
Mean	.72	.72	.73	.73
Optimal BW	3.53	3.57	5.32	5.32
N	15743	15701	23240	23240

Note: Table A.6 presents non-parametric estimations of the effect of retention and accelerated instruction on each outcome. Columns 1 and 3 include controls only for cohort. Columns 2 and 4 include controls for student and school characteristics and cohort fixed effects. Reclassification is measured as whether the student has been reclassified by each grade. The Optimal BW rows show the bandwidth used for each estimation. The Mean rows show the mean of each outcome among students who passed the third attempt of the reading test and their scores are within the optimal bandwidth. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, and had valid scores for each test administration of the TAKS assessment. The number of observations for each outcome varies due to missing observations and the optimal bandwidth chosen. Standard errors, shown in parenthesis, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.

Table A.7: Effect of Third-Grade Retention on all Outcomes Restricting the Sample to Students with Information on Earnings at Ages 20-22

	(1) ITT	(2) TOT
Panel I: Academic Achievement		
Reading Grade 4	0.328***	0.940***
	(0.044)	(0.100)
Math Grade 4	0.282***	0.885***
	(0.045)	(0.107)
Writing Grade 4	0.195***	0.566***
	(0.025)	(0.063)
Reading Grade 5	0.300***	0.849***
	(0.047)	(0.096)
Math Grade 5	0.266***	0.790***
	(0.047)	(0.100)
Panel II: Reclassification		
Yrs classified as EL	-0.205*	-0.497
	(0.117)	(0.305)
Ever reclassified	0.065***	0.096**
	(0.024)	(0.046)
Reclassified by Grade 4	0.049***	0.159***
	(0.018)	(0.040)
Reclassified by Grade 5	0.076***	0.237***
	(0.021)	(0.053)
Rec1assified by 1 year later	-0.004	-0.021
	(0.014)	(0.037)
Rec1assified by 2 years later	0.049**	0.150***
	(0.020)	(0.051)
Panel III: High school completion		
Ever Graduated	-0.011	-0.025
	(0.017)	(0.043)
Graduated 26 plan	0.008	0.021
	(0.020)	(0.051)
On time graduation 4 years in high schoolhs	-0.021	-0.042
	(0.019)	(0.048)
On time graduation (Cond. on retention)	0.065***	0.192***

Continued on next page

	(1) ITT	(2) TOT
	(0.021)	(0.050)
Dropout	0.005	0.018
	(0.013)	(0.033)
Panel IV: Enrollment in higher education		
Enr. any institution (17-20 years old)	-0.007	-0.012
	(0.018)	(0.050)
Enr. community college (17-20 years old)	-0.006	-0.013
	(0.018)	(0.048)
Enr. any institution (9-12 years post 3rd grade)	-0.011	-0.030
	(0.019)	(0.049)
Enr. community college (9-12 years post 3rd grade)	-0.014	-0.043
	(0.018)	(0.047)
Panel V: Earnings		
Av Quarterly Earnings	0.014	-0.040
	(0.073)	(0.167)
Av Quarterly Earnings(not enrolled in school)	-0.039	-0.088
	(0.053)	(0.133)

Note: Table A.7 presents point estimates of non-parametric estimations of the effects of third-grade retention and accelerated instruction on each outcomes restricting the sample to students who had information on earnings at ages 20 to 22. The sample includes first-time third-graders from 2002-2003 through 2008-2009, who were enrolled in the following academic year, did not receive special education services, had valid scores for each test administration of the TAKS assessment, and had information on earnings. Standard errors, shown in parentheses, are clustered at the school level.*** p<0.01, ** p<0.05, * p<0.10.