Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1557

# Remediation or Roadblock? The Effects of Failing Virginia's Accountability Exams

Arnav Dharmagadda
University of Virginia

Virginia's Standards of Learning (SOL) system requires students to pass end-of-course exams to earn verified credits needed for high school graduation. Scoring below the passing cutoff may create a barrier to completing a high school degree, but it also triggers state-mandated remediation, which may include additional instruction and tutoring. Because these channels push in opposite directions, the net effect on student outcomes is theoretically ambiguous. This paper estimates causal effects using detailed data from the Virginia Longitudinal Data System (VLDS) and a regression discontinuity design, comparing students who score just below the passing cutoff to those just above. Whether students marginally pass or fail course-level exams has no detectable effects on high school graduation. For Algebra I, scoring below the cutoff raises first-attempt pass rates on subsequent Geometry and Algebra II exams by approximately 2 percentage points (around a 5 percent increase) and increases postsecondary enrollment within 16 months of high school graduation. These gains are consistent with remediation producing meaningful skill improvement. For English Reading, which is generally taken later in high school, scoring below the cutoff produces no detectable effects on subsequent attainment. There is no evidence that marginally failing reduces attendance, narrows course-taking, or discourages college preparation — the patterns one would expect from stigma or disengagement. Yet, scoring below the expedited retake cutoff in Algebra I, causing students to face a mandatory delay before retaking the exam, reduces high school graduation.

VERSION: August 2026

Suggested citation: Dharmagadda, Arnav. (2026). Remediation or Roadblock? The Effects of Failing Virginia's Accountability Exams. (EdWorkingPaper: 26-1557). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/4f0e-yw52

# Remediation or Roadblock?
# The Effects of Failing Virginia's Accountability Exams*

Arnav Dharmagadda†

August 9, 2026

**Abstract**

Virginia's Standards of Learning (SOL) system requires students to pass end-of-course exams to earn verified credits needed for high school graduation. Scoring below the passing cutoff may create a barrier to completing a high school degree, but it also triggers state-mandated remediation, which may include additional instruction and tutoring. Because these channels push in opposite directions, the net effect on student outcomes is theoretically ambiguous. This paper estimates causal effects using detailed data from the Virginia Longitudinal Data System (VLDS) and a regression discontinuity design, comparing students who score just below the passing cutoff to those just above. Whether students marginally pass or fail course-level exams has no detectable effects on high school graduation. For Algebra I, scoring below the cutoff raises first-attempt pass rates on subsequent Geometry and Algebra II exams by approximately 2 percentage points (around a 5 percent increase) and increases postsecondary enrollment within 16 months of high school graduation. These gains are consistent with remediation producing meaningful skill improvement. For English Reading, which is generally taken later in high school, scoring below the cutoff produces no detectable effects on subsequent attainment. There is no evidence that marginally failing reduces attendance, narrows course-taking, or discourages college preparation — the patterns one would expect from stigma or disengagement. Yet, scoring below the expedited retake cutoff in Algebra I, causing students to face a mandatory delay before retaking the exam, reduces high school graduation.

*For helpful comments and discussions, I thank Sarah Turner, Leora Friedberg, Diego Briones, Michael Ricks, Andrew Simon, Daniel Tannenbaum, Brenden Timpe, Rebecca Jack, and participants at the Nebraska Labor Summit and Workshop. For data access and support, I thank Ron Parrish of the Virginia Department of Education and Tod Massa of the State Council of Higher Education for Virginia.

†University of Virginia. qbd8dd@virginia.edu

# 1 Introduction

Each year, around 90,000 Virginia public school students take the Algebra I End-of-Course Exam, and around 26,000, or 29%, do not pass on their first attempt. The Algebra I exam and its English counterpart form part of the state's high school Standards of Learning (SOL) assessment framework, a set of end-of-course exams that are a core part of graduation requirements. Failing an SOL exam, defined as scoring below 400 on the 600-point scale, poses a direct barrier to on-time graduation. But receiving a failing score also provides a potentially useful snapshot of a student's ability for schools, teachers, parents, and students themselves, and it triggers an institutional response: Virginia's state code requires that students identified as low performers based on their SOL scores receive remediation, such as additional instructional time, personal tutoring, or specialized coursework. The bundle of responses to failing can have opposing effects on student outcomes, making the net effect of failing unclear. This paper investigates how failing the Algebra I SOL exam and its English counterpart affects subsequent high school performance, high school completion, and postsecondary attainment.

The two opposing forces at work when a student fails, a barrier to graduation on one side, mandated remediation on the other, have been studied largely in isolation. Existing research on failing high school accountability exams focuses on traditional exit exams, which students must pass to graduate, and documents null-to-negative effects of failing (Reardon et al., 2010; Clark and Martorell, 2014; Papay et al., 2010). Previous work studying remediation based on exam performance at the elementary and middle school level suggests it can improve future achievement (Schwerdt et al., 2017; Figlio and Ozek, 2024). Closest to the present study, Özek (2025) shows that, in one district, a Florida high school remedial reading course policy based on test scores reduces postsecondary enrollment and degree attainment, potentially driven by the crowding out of advanced coursework and electives. Virginia's SOL framework presents a policy context with course-specific tests, multiple pathways to attain graduation requirements, and mandated remediation based on exam performance at the high school level. Unlike test-based retention policies, Virginia students can remediate and retest within the same year if they score within 25 points of passing, providing a policy environment that isolates remediation effects from large delay costs.

The Virginia Longitudinal Data System (VLDS), a statewide panel dataset with the detailed information on public school student enrollment and performance, provides the necessary detail for comparing short- and long-term outcomes of students who fail and students who pass an SOL exam. On average, students who fail have lower attendance, graduation, and college enrollment than those who pass, but these gaps largely reflect differences in the students themselves rather than the effect of failing. Students who score below the cutoff have weaker academic preparation than those who score above it, so a simple comparison of outcomes confounds the effect of failing with the effect of lower ability. A regression discontinuity design addresses this problem by comparing students who score just below the passing cutoff to those just above, where differences in underlying ability are negligible. Any discontinuous jump in the trend at 400, the passing threshold, is a causal effect of failing, bundled with the associated student and institutional responses. The effect is a local average treatment effect (LATE), isolating the marginal effect of failing from the effect of underlying differences in ability and behavior.

Nearly all Virginia public school students take the Algebra I and English Reading SOL exams — Algebra I typically in middle school or early high school, English Reading in 11th grade — making these two exams the natural focus of the regression discontinuity analysis. Scoring below the passing cutoff does not have a detectable effect on high school graduation for either exam. The subjects diverge on other margins. Scoring below the cutoff on Algebra I increases enrollment in Geometry and Algebra II by 1 – 3 percent and improves performance on the Geometry and Algebra II SOL exams by roughly 5 percent, consistent with remediation producing real skill gains. In the longer term, scoring below the cutoff in

Algebra I increases postsecondary enrollment by about 2 percent. For English Reading, taken late enough that little high school curriculum remains, there are no detectable downstream effects. Across both subjects, there is no evidence that marginally failing reduces attendance or college preparation, measured in the form of college entrance exam participation and AP test participation. These patterns are inconsistent with negative stigma or disengagement from scoring below the passing cutoff.

Students who score below a 375 on their first attempt, however, see negative effects on long-run outcomes. Virginia prevents these students from retaking in the same testing window, introducing a delay costs. For students at the 375 threshold in Algebra I, scoring below the cutoff reduced enrollment in Algebra II and reduced high school graduation. Virginia provides evidence that remediation initiated by exams can lead to skill gain, but understanding the specific aspects of remediation and other responses to failing that drive differential results across subjects and thresholds remains an open question.

## 2 Policy Context

### 2.1 What Do Accountability Systems Do?

Accountability systems encourage schools to improve student performance by measuring it against established standards and attaching meaningful consequences contingent on it. Empirical evidence shows that introducing an accountability system to a state can improve student achievement (Hanushek and Raymond, 2005; Dee and Jacob, 2011), though these gains are not necessarily equally distributed across all students (Neal and Schanzenbach, 2010).

In addition to the incentive effect on schools, accountability systems provide information to students, parents, teachers, and schools about a student's ability level relative to proficiency standards. This *labeling effect* of test scores can influence long-term student outcomes, including subsequent course performance, graduation, and postsecondary outcomes, depending on how stakeholders respond. Students can have personal reactions to their scores that influence future decision-making. State policy can block students with low scores on certain tests, such as exit exams, from graduating. Schools and teachers can target at-risk students identified based on failing test scores with remediation, which can affect achievement.

Much of the existing research on labeling effects at the high school level studies how failing an exit exam affects student outcomes and finds null to negative results. There was no detectable effect on course-taking or graduation for students who fail by a small margin in California (Reardon et al., 2010), and in Massachusetts, failing the exit exam left the general student population unaffected while reducing graduation rates for low-income urban students (Papay et al., 2010). Even where students who marginally pass are substantially more likely to receive a diploma, as in Clark and Martorell (2014), the earnings returns to passing are negligible, suggesting that the diploma itself conveys a limited additional signal beyond what marginal students already possess.

Other research at the primary and middle school levels in states where a low test score initiates remediation shows that failing can have positive effects. Figlio and Ozek (2024) find that failing a Florida middle school achievement test triggers tracking into lower-performing peer groups and reduces access to advanced coursework, but that remediation also meaningfully improves future test scores. Schwerdt et al. (2017) similarly find that Florida's third-grade reading retention policy improves short-term achievement and high school GPA. At the high school level, Cortes et al. (2014) find that assigning students with low eighth-grade math scores to a double-dose algebra course raised math performance, high school graduation, and college enrollment. Özek (2025), by contrast, finds that in one Florida district a policy requiring low-performing students to take an additional remedial reading course lowers advanced course-taking and postsecondary attainment. Looking at grade retention policies for older students, there is evidence consistent with sizable delay costs: Manacorda (2012) finds retention of middle school students in Uruguay increased dropout likelihood, while Jacob and Lefgren (2009) show that retaining eighth-graders in Chicago increased dropout rates.

Virginia's end-of-course framework provides a policy setting where mandatory remediation is paired with minimal delay costs. A student who marginally fails is neither held back a grade nor required to repeat the course, and after remediation those scoring at or above 375 may retake within the same testing window. The estimate at the 400 cutoff therefore largely reflects the remediation response without the substantial delay costs that accompany failing in retention systems. That delay cost, combined with remediation, resurfaces at the 375 threshold, where the expedited retake is unavailable.

### 2.2 Virginia's Standards of Learning (SOL)

Virginia lawmakers adopted the Standards of Learning (SOL) curriculum for math, English & language arts (ELA), science, and social studies education in 1995. In the spring of 1998, Virginia public school students took the first administration of the SOL exams (Yecke, 1999). After the No Child Left Behind Act (NCLB) passed in 2002, Virginia used the SOLs as its official assessment for federal accountability requirements. NCLB mandated that schools that failed to demonstrate adequate performance on standardized tests would eventually have to allow students to transfer, provide additional support to students, and take corrective action (including restructuring and state takeover) (Applequist, 2008; McGuinn, 2016; U.S. Department of Education, 2007). Penalties under NCLB established incentives for schools to improve student scores. The SOL assessment remained the state's official exam when the Every Student Succeeds Act (ESSA), which weakened accountability requirements, replaced NCLB in 2015 (Virginia Department of Education, 2025; McGuinn, 2016). By 2025, Virginia was one of only 6 states that still required standardized tests as part of graduation requirements, reflecting declining public support for high-stakes testing (FairTest, 2017).

The SOL accountability system at the high school level has an end-of-course model that other states have increasingly adopted to replace exit exams (Zabala et al., 2008). The high school end-of-course exams are directly tied to the content of a specific class, such as Algebra I or Biology, whereas between third and eighth grade, the annual SOL exams are comprehensive across general subjects like math and reading. The structure of each SOL exam is shared: a student must score 400 on a 600-point scale to pass and earn verified credit for any SOL (Virginia Department of Education, 2021, n.d.b). Every SOL exam, except the SOL Writing Exam, is computer-based and computer-graded. The SOL Writing Exam is completed online, but it is graded by a person. Students must answer multiple-choice or technology-enhanced questions, such as matching, short answer, and drag-and-drop questions (Virginia Department of Education, n.d.a).

Students have incentives to perform well on SOL exams because passing earns a verified credit, and they must earn a set number of verified credits to receive a diploma (Virginia Department of Education, n.d.c, 2021).[1] Students need to earn at least 2 English verified credits, 1 Math verified credit, 1 Science verified credit, and 1 History/Social Sciences verified credit to graduate. Prior to the Class of 2022, students needed to earn an additional verified credit in a subject of their choice and had the option to earn additional verified credits to qualify for an advanced diploma. Multiple SOL exams can fulfill a verified credit requirement. For example, a student who must earn 1 verified credit in Math can obtain a verified credit in Algebra I, Geometry, or Algebra II. Alternate exams, such as SAT, ACT, and AP exams, can substitute for some verified credits.

State regulations mandate that all students who fail an end-of-course SOL required for a

[1]Under select criteria, some students will be eligible to earn locally-awarded verified credits (LAVCs), which allow students to fulfill graduation requirements without passing an SOL exam. The student must (1) pass the associated high school course, (2) score within a 375-399 after taking the SOL exam at least twice, and (3) complete a locally-approved process. LAVCs were not awarded for math and reading SOLs until May 2018 (McKenna, 2018).

verified credit attend remediation. Specifics of the remediation, such as length, are not set by the Virginia Department of Education; rather, they are largely controlled by individual schools and districts, following guidance from the Virginia Board of Education (Virginia Department of Education, 2024, 2021). After remediation, students who fail an SOL exam have multiple options to earn their verified credit. Some students retake the test. Students who score between 375 and 399, who passed the associated course, and who completed a remediation program are eligible for an *expedited retake* within the same testing window. Students who score below 375 must wait until the next testing window to retest. Other students will take a different SOL exam to fulfill the content area requirement. For example, a student who fails Algebra I may continue to the Geometry SOL and earn their verified credit that way. Students may also satisfy the verified credit requirement through approved, alternative tests.[2]

### 2.3 Competing Effects at the Pass-Fail Cutoff

An organizing framework helps conceptualize the effect of remediation and other responses to failing an SOL exam. Consider the effect of failing, $\tau$, on a student outcome as being composed of three individual channels – remediation ($r$), stigma ($\sigma$), and sequencing ($\delta$) – which are functions of ability, $a$, and failing. Let ability vary smoothly with the test score, and let $\bar{a} \equiv a(c)$ denote the ability of students at the cutoff, $c$. Assume continuity of ability in the running variable. As scores approach the cutoff from either side, and ability converges to the common value $\bar{a}$. Let $D_i \in \{0, 1\}$, where $D_i = 1$ means a student failed the SOL exam, and $D_i = 0$ means a student passed the SOL exam. The effect $\tau$ is then the difference between the fail-side limit ($D_i = 1$, scores just below $c$) and the pass-side limit ($D_i = 0$, scores just above $c$), both evaluated at the ability $\bar{a}$.

[2] See Figure A1 for an overview of the verified credit regime.

$$\tau = \underbrace{r(\bar{a}, D_i = 1) - r(\bar{a}, D_i = 0)}_{\text{remediation gain } \geq 0} + \underbrace{\sigma(\bar{a}, D_i = 1) - \sigma(\bar{a}, D_i = 0)}_{\text{stigma cost } \lesseqgtr 0} + \underbrace{\delta(\bar{a}, D_i = 1) - \delta(\bar{a}, D_i = 0)}_{\text{sequencing cost } \lesseqgtr 0} \tag{1}$$

First, the remediation channel refers to additional support, resources, and learning that occur because a student has failed. Students may be enrolled in formal or informal remediation programs, or they may self-study for retakes (Figlio and Ozek, 2024; Schwerdt et al., 2017). Students may have additional time to acquire the necessary human capital. At the extreme end, students may retake classes. Assume that the remediation channel is weakly positive for all students, $r(a, D_i) \geq 0$, but the returns to remediation are diminishing with the ability $\frac{\partial r}{\partial a} < 0$. If remediation is only delivered to students who fail the exam, or at least more remediation is delivered to students who fail the exam, then $r(\bar{a}, D_i = 1) - r(\bar{a}, D_i = 0) \geq 0$.

Second, when a student fails an exam, the failing score creates a negative signal for the student themselves as well as their teachers and school administrators. The negative signal may impair self-perception of ability, causing disengagement in the classroom and discouragement (Reardon et al., 2010). The negative signal may alternatively increase student motivation if they interpret the failing score as an indication that they need to invest more effort. Outside a remediation program, schools or teachers may change how they allocate attention and resources to students they view as failing. Some teachers may devote more attention to marginal students who are most likely to pass with additional support (Neal and Schanzenbach, 2010), while other teachers may reduce their investment in students who fail if the teachers believe increased attention will not translate into improved achievement. The stigma responses will depend on how students and teachers interpret their score as a signal, making the sign ambiguous.

Finally, there are sequencing effects associated with failing an exam. Remediation, either informal or formal instruction, can displace other coursework and academic opportunities.

Students and schools may choose to alter student schedules in response to failing. Students may take coursework more suited for their ability level in response to failing, which could benefit them, or failing may displace coursework that would have been beneficial. Students may be exposed to different peers in the new coursework, which could lead to positive or negative peer effects (Figlio and Ozek, 2024; Schwerdt et al., 2017). The net sign on the sequencing effect is ambiguous.

Because the remediation effect is positively signed, but stigma and sequencing effects are ambiguous, the sign of $\tau$ is ambiguous a priori.

## 3 Data and Descriptives

The Virginia Longitudinal Data System (VLDS) is a federated data system that links databases from Virginia state agencies, including the Virginia Department of Education (VDOE) and the State Council of Higher Education for Virginia (SCHEV). The analysis uses VDOE student records from the graduating cohorts of 2010 – 2019, which avoid pandemic-induced changes to testing schedules and allow sufficient time to capture longer-run outcomes. The data represent all students who attended a Virginia public school. Each graduating cohort includes around 90,000 students, for a total of approximately 900,000 individual students. The VLDS contains information on course enrollment and grades, SOL exam administrations and scores, attendance, high school graduation outcomes, and college entrance exam administrations and scores. The VLDS includes data from the National Student Clearinghouse (NSC) on post-secondary enrollment and contains a rich set of demographic identifiers, including eligibility classifications for free and reduced-price lunch.

The analysis focuses on Algebra I and English SOL exams, which are nearly universally taken by Virginia students between 2010 and 2019. During the sample period, Virginia implemented revised standards for Algebra I and English, resulting in new exams designed to assess proficiency against these benchmarks. Different exams within the same subject are designated with distinct test codes. The year the student takes the test largely determines the test code that a student receives.[3] The distribution of test scores differs markedly across test codes. The English test code 109 is harder to pass than 30 or 209, while the Algebra I test code 20 is easier to pass relative to 120 and 220.[4] The scaled score 400 represents different ability thresholds across test codes. Separating the analysis by test code avoids conflating fundamentally different treatment and control groups.[5]

Table 1 presents summary statistics for the analysis sample. The average student takes Algebra I in 8th or 9th grade, and the English exam is typically administered in the 11th grade. The samples are large — 363,705 students for Algebra I and 539,088 for English — and reflect Virginia's racial and socioeconomic diversity: around 36% of students qualify for free and reduced-price lunch, and slightly under half are non-White. Failure rates on the first attempt are non-trivial: 29% of students fail Algebra I, and 17% fail English. These failures are not evenly distributed. Black, Hispanic, and economically disadvantaged students fail at substantially higher rates, which motivates the heterogeneity analysis in Section 5.

The same table's outcome-mean panels show that the raw differences between those who pass and those who fail are large across every dimension: students who fail Algebra I have 44% lower enrollment in Algebra II, 3% lower attendance in subsequent years, 11–12% lower graduation rates, and 38–47% lower post-secondary enrollment rates. The gaps cannot be interpreted causally — students who pass will, on average, have higher ability than students who fail. The empirical strategy in Section 4 isolates the effect of failing from the underlying ability difference.

[3]In Algebra I, test code 20 transitioned to test code 120 beginning in 2011, and test code 120 transitioned to test code 220 in 2018. In English, test code 30 transitioned to test code 109 in 2012, and test code 109 transitioned to test code 209 in 2019.

[4]Figure A2 shows the distribution of performance by test code.

[5]Within test codes, students will be administered different questions based on the test form they receive. Figure A3 shows that year-to-year score distributions within test codes look similar, mitigating the risk of substantial test form differences in student scores.

The majority of students who fail — 62% in Algebra I and 67% in English — retake within the same school year, consistent with schools directing students toward the state-mandated expedited retake process.[6] The abundance of same-year retakes informs the interpretation of the estimated effect, $\tau$, in the paper. That said, a plurality of students who fail their first attempt never pass despite having essentially unlimited retake opportunities, a pattern that may reflect the availability of alternative pathways to graduation.

## 4 Empirical Strategy

A regression discontinuity design identifies the causal impact of the treatment, defined as scoring below the passing cutoff of 400 on an end-of-course SOL exam. The passing cutoff creates a sharp, discontinuous change in treatment status at a score of 400. Under the continuity assumption — that students' potential outcomes evolve smoothly through the cutoff absent treatment — any discontinuous jump in outcomes at the threshold reflects the causal effect of failing.

The estimating equation is:

$$Y_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta \left( \widetilde{S}_i \times D_i \right) + \gamma_c + \varepsilon_i, \quad \text{for Score}_i \in [400 - h, 400 + h] \tag{2}$$

where $D_i = \mathbf{1}[\text{Score}_i < 400]$ is an indicator for failing the SOL exam, $Y_i$ is an outcome, and $\widetilde{S}_i = \text{Score}_i - 400$ is the running variable centered at the cutoff. The interaction term allows the linear slope to differ on each side of the threshold, and $\tau$ captures the causal effect of failing the SOL exam on future outcomes. $\gamma_c$ denotes cohort fixed effects,[7] and standard

[6] See Table A1 for additional statistics on what happens after a first-attempt failure.

[7] Cohort fixed effects address a compositional imbalance at the threshold. Because the score mass points vary across years, as shown in Figure A4, the share of students from each cohort can differ discontinuously just

errors are clustered at the school-level since schools determine remediation responses. In secondary specifications, $D_i = \mathbf{1}[\text{Score}_i < 375]$, an indicator for failing to meet eligibility for an expedited retake, with the running variable re-centered at 375. To complement the analysis, the main specification is re-estimated adding a quadratic term for the running variable, serving as a robustness check on the sensitivity of the estimates to functional form assumptions.

The continuity assumption underlying regression discontinuity requires that students cannot manipulate their scores around the threshold in ways correlated with potential outcomes. A standard check for manipulation looks for the smooth density of scores around the threshold. The density of scores is not smooth through the cutoff. For example, in Algebra I, the scores 397 and 401 are uncommon relative to 399 and 402.[8] This imbalance is a mechanical artifact of the test itself, and it is not student sorting. The conversion of a raw score on a multiple choice assessment to a 600 point scale produces unevenly populated mass points throughout the entire score distribution, not just around the cutoff.[9] Both Algebra I and English are computer-graded, and students are unaware of the exact raw score needed to pass, making sorting implausible and supporting the continuity assumption.

To make the estimated coefficients less sensitive to unevenly populated mass points, the analysis estimates Equation 2 over wide, symmetric bandwidths of $h \in \{10, 15, 20, 25\}$ points, with $h = 20$ — the window $[380, 420]$ — as the preferred specification. The wide bandwidths ensure that no single score value exerts substantial leverage on the estimated slopes or the cutoff intercept. Providing a menu of bandwidths and checking for consistency in the direction

below and just above the cutoff. There exist cohort-level differences in outcomes, potentially driven by labor market conditions or the policy environment. For example, a strong labor market for high school graduates without a college degree will influence college enrollment. If cohort distributions evolved smoothly through the cutoff, $\tau$ would represent a weighted average of the effect across cohorts. But with a discontinuous shift in cohort composition, cohort-level differences in outcomes can contaminate $\tau$. Including cohort fixed effects absorbs these cohort-level mean differences in outcomes, ensuring that identification comes from within-cohort variation rather than across-cohort differences.

[8] See Figure A5 for the distribution of Algebra I scores near the cutoff.

[9] See Figure A5 for the full distribution.

and magnitude of the coefficients further serve as robustness checks. The widest bandwidth also respects a natural institutional boundary: the left edge of the $[375, 425]$ window coincides with the expedited-retake eligibility threshold at 375, so the failing students within every bandwidth face identical retake abilities. The mean squared error optimal bandwidth selector is not directly applicable, since it assumes a continuous running variable.

The continuity assumption cannot be tested directly because the full measure of student ability is unobserved. But observed predetermined characteristics — gender, race and ethnicity indicators, and economic disadvantage — cannot be affected by failing, so their smooth evolution through the cutoff offers supportive evidence (Lee and Lemieux, 2010). None of the six covariates shows a meaningful discontinuity at the passing cutoff, individually or jointly: at the preferred bandwidth, no covariate is significant at the 0.05 level, and a joint test fails to reject that all six discontinuities are zero. The same balance holds for the preferred bandwidth at the 375 expedited-retake threshold.[10]

# 5 Results and Discussion

## 5.1 Downstream Math Enrollment and Performance

Estimating Equation 2 with future math course enrollment and math SOL performance as dependent variables identifies the impact of Algebra I failure and resulting remediation on downstream math coursework. Failing Algebra I and undergoing the resulting remediation improves students' subsequent trajectory through the high school math sequence.

Scoring below the cutoff on the Algebra I SOL increases enrollment in subsequent math courses. Table 2 shows that failing Algebra I and the resulting remediation increases the likelihood of enrolling in Geometry by 0.8 percentage points, significant at the 0.10 level, and

[10]See the Appendix for covariate balance plots (Figures A7 and A8) and formal balance tests (Tables A2, A3, A4, and A5).

increases the likelihood of enrolling in Algebra II by 1.7 percentage points or 2.8%. RD plots corresponding to this specification and those discussed below are presented in Appendix C.

Scoring below the cutoff on the Algebra I SOL increases performance on the Geometry and Algebra II SOL exams taken in later years. Table 2 shows that students are around 5.3% more likely to pass Geometry and 5.6% more likely to pass Algebra II, statistically significant and economically meaningful effects. The stability of point estimates when examining the effect on performance conditional on test-taking suggests there is no strong selection into test-taking. In terms of raw point totals, failing causes 2.6 points of improvement in Geometry and 2.3 points of improvement in Algebra II. The effects on performance are broadly similar in direction and magnitude for Black & Hispanic and economically disadvantaged students, but the effect is larger for male students. The direction of the effects are consistent across bandwidth specification.

The direction of the enrollment and performance results for both exams is robust to the alternative quadratic specification. The enrollment effect in Geometry loses precision and the performance effects are larger in magnitude.[11] The linear model remains the preferred specification because it is less vulnerable to overfitting.

### 5.2 High School Completion and Post-Secondary Enrollment

Scoring below the cutoff in Algebra or English does not substantially affect high school graduation or high school completion, which includes GED students. Table 3 shows a precise null result. In Algebra I, the analysis rules out effects on graduation below -0.4 and above 0.8 percentage points. In English, the analysis rules out effects below -0.4 and above 0.4 percentage points. Since graduation rates are very high, at around 95%, these ranges represent small, economically insignificant effects.

[11] See Table A10.

There are similar null effects of failing the Algebra I or English SOL on dropping out, but these effects are estimated with less precision. The analysis rules out effects below -0.2 and above 0.6 percentage points in Algebra I and below -0.4 and above 0.4 percentage points in English. These small point estimates translate into a wide range of percent effects due to the low 3% baseline dropout rate. The low baseline dropout rate limits the variance in the outcome, reducing the precision of the estimates.

The same table presents the postsecondary results. There is suggestive evidence that scoring below the cutoff on the Algebra I SOL exam increases postsecondary enrollment within 16 months of graduation by 1 percentage point (2%). This positive effect, significant at the 0.10 level, is robust to bandwidth specification. The analysis lacks sufficient statistical power to identify separate effects on 2-year and 4-year college enrollment. There is no evidence that failing reading affects college enrollment.

The effects on high school completion and post-secondary enrollment are consistent with a quadratic specification.[12]

### 5.3 Potential Channels

To understand whether students respond to failure by altering engagement patterns in school, the analysis first examines the effect on attendance. Across both English and Algebra I, failing the SOL exam does not affect attendance in subsequent years; Table 4 shows a precise null effect. There is also no evidence that students disengage from college preparation. Failing either exam does not have a statistically significant effect on taking the college entrance exam (ACT/SAT). Students who fail Algebra I are 4.4% more likely to take a college-credit-awarding AP exam, potentially a consequence of skill gain in subsequent math classes. The result is significant at the 0.10 level and generally robust to bandwidth

[12]See Table A11 and Table A12.

selection. The analysis also documents increased enrollment in career and technical education (CTE) coursework after failing Algebra I and increased completion of a CTE competency after failing English. The course-taking results are the opposite of what Özek (2025) finds in his analysis of a Florida school district remedial course policy, a difference potentially explained by the different structure and timing between the two systems.

Failing English increases new disability classifications by 0.2 percentage points, or 22.2% off of a 0.9% baseline rate. The direction of the result is consistent across bandwidths and significant at the 10% level. Prior research has documented special needs classifications in response to failing an exam (Jacob, 2005). There is no significant effect of scoring below the cutoff in Algebra I on new disability classifications.

### 5.4 Examining Effects at the 375 Threshold

Virginia's institutional structure creates a separate treatment at the 375 threshold, distinct from passing or failing at the 400 threshold. Students who score above 375 are eligible for an expedited retake within the same exam window, while those below 375 must wait until the next testing administration. The analysis at this threshold focuses on a different treatment applied to a different student population.

The downstream math response reverses: students who fall below 375 on the Algebra I SOL are 1.7 percentage points less likely to enroll in Algebra II (significant at the 0.05 level), whereas marginally failing at 400 raised Algebra II take-up by a similar margin. This negative effect is stronger for Black and Hispanic, economically disadvantaged, and male students. There is no evidence that scoring below 375 affects performance on subsequent math SOL exams, not conditional on taking the exam.

Table 5 reports the high school completion and postsecondary results at this threshold. Scoring below 375 on the Algebra I SOL reduces high school graduation by 1.5 percentage points (significant at the 0.05 level), representing a 1.7% decrease. The result, which contrasts with the precise null at the 400 cutoff, is robust to bandwidth choices, though it loses precision at narrower bandwidths. Men have a stronger response to scoring below a 375, reducing high school graduation by 3 percentage points. There is no statistically significant effect in English. There is no precise estimate of the effect on post-secondary enrollment for the general population from scoring below a 375 on either exam. Scoring below a 375 on the English exam reduced enrollment in 4-year colleges by 1.8 percentage points for Black and Hispanic students and 1.6 percentage points for economically disadvantaged students, representing declines of 12.7% and 15.1%, respectively. Scoring below a 375 in Algebra I and English reduced CTE family enrollment by 1.4 percentage points and 1.6 percentage points,[13] a displacement of career and technical coursework that mirrors the crowd-out channel Özek (2025) identifies as a driver of the adverse effects of high school remediation.

Delay costs for a retake are a key difference between the 375 and 400 thresholds. Students who score below 375 are far less likely to retake the SOL exam in the same year. Conditional on retaking the exam, in Algebra I, 27.3% of students scoring below a 375 on their first attempt retook within the exam within the same year. For students who scored between 375 and 400, it was 95.6%.

## 6 Conclusion

The passing cutoff on the SOL is a performance standard. When a student falls short on the first attempt, does their achievement trajectory change — either toward achievement gains or toward setbacks? Scoring below the Algebra I cutoff produced measurable gains in later math performance and suggestive evidence of higher postsecondary enrollment; scoring below the English cutoff produced no detectable effects on later performance or postsecondary

[13] See Table A7 for full mechanisms results.

enrollment. Neither exam affected high school graduation. The regression discontinuity design has two important limitations. First, it identifies effects local to students near each threshold. Second, the design identifies the effect of failing and the bundle of responses associated with it; it cannot distinguish between the effect of remediation and that of other stigma and sequencing responses.

While the effects at the passing threshold are positive or null, the lower 375 cutoff, where failing forces a longer wait before retaking, carries more penalty than benefit: in Algebra I, graduation rates fall, and downstream math gains disappear. The two thresholds distinguish different marginal students, so the contrast cannot be interpreted as the pure effect of the retake delay. But it raises the question of whether the pace of remediation and retakes may affect how students recover from a failing score, an issue left for future investigation.

The estimates capture everything a failing score sets in motion — school-directed remediation, shifts in how teachers allocate attention, and students' and families' own responses to the signal — without observing which of these drives the net effect. Schools have wide latitude in how they remediate, so the same failing score can trigger very different institutional responses across districts. Prior work has shown that providing information to parents increases student effort and achievement (Bergman, 2021). Opening this black box by linking outcomes to details of how remediation is actually delivered and to how students, teachers, and parents interpret and act on the score is the natural next step. It would show which responses drive the gains and which dimensions of remediation (its timing, intensity, and content) determine whether skill gains persist.

Whether the observed effects persist in contexts beyond Virginia's forgiving system, where multiple pathways satisfy exam-based graduation requirements and most students who fail retake within the same year, remains an open question. Even so, as states rethink the design of their accountability systems, Virginia provides evidence that behavioral and institutional responses to failing course-specific accountability exams can improve student achievement without affecting graduation.

# 7 Main Tables

Table 1. Sample Demographics and Outcome Variable Means: Algebra I and English

| | Algebra I (TC120) | | | English (TC109) | | |
|---|---|---|---|---|---|---|
| | Full Sample (1) | Failed (<400) (2) | BW [380, 420] (3) | Full Sample (4) | Failed (<400) (5) | BW [380, 420] (6) |
| *Panel A: Demographics* | | | | | | |
| Students ($N$) | 363,705 | 103,914 | 118,335 | 539,088 | 90,335 | 119,545 |
| Female (%) | 49.5 | 43.7 | 48.7 | 49.7 | 44.5 | 47.7 |
| Black (%) | 23.4 | 34.5 | 29.4 | 22.6 | 39.8 | 34.9 |
| White (%) | 53.8 | 44.0 | 49.5 | 54.5 | 34.8 | 41.6 |
| Hispanic (%) | 11.8 | 14.7 | 13.1 | 11.6 | 17.8 | 14.7 |
| Asian (%) | 6.2 | 2.2 | 3.2 | 6.7 | 4.2 | 4.6 |
| Economically Disadvantaged (%) | 37.6 | 54.6 | 45.6 | 34.6 | 61.8 | 51.2 |
| Mean grade when test taken | 8.56 | 9.02 | 8.80 | 11.05 | 11.15 | 11.07 |
| Mean first SOL score | 427.8 | 374.9 | 401.6 | 439.9 | 372.3 | 403.2 |
| Passed on first attempt (%) | 71.2 | 0.0 | 56.1 | 82.8 | 0.0 | 62.6 |
| *Panel B: Downstream Academic Outcomes (Algebra I only)* | | | | | | |
| Took Algebra II Course | 0.692 | 0.389 | 0.614 | — | — | — |
| | (0.462) | (0.488) | (0.487) | | | |
| Took Geometry Course | 0.869 | 0.787 | 0.876 | — | — | — |
| | (0.337) | (0.409) | (0.329) | | | |
| Algebra II Pass | 0.821 | 0.555 | 0.660 | — | — | — |
| | (0.383) | (0.497) | (0.474) | | | |
| Geometry Pass | 0.731 | 0.378 | 0.570 | — | — | — |
| | (0.444) | (0.485) | (0.495) | | | |
| *Panel C: Attendance* | | | | | | |
| Attendance Rate, Year 1 | 0.948 | 0.922 | 0.942 | 0.930 | 0.904 | 0.918 |
| | (0.072) | (0.097) | (0.071) | (0.086) | (0.118) | (0.097) |
| Attendance Rate, Year 2 (Algebra I only) | 0.942 | 0.913 | 0.933 | — | — | — |
| | (0.082) | (0.111) | (0.084) | | | |
| *Panel D: High School Completion* | | | | | | |
| HS Graduate | 0.936 | 0.835 | 0.943 | 0.960 | 0.847 | 0.955 |
| | (0.244) | (0.371) | (0.231) | (0.197) | (0.360) | (0.207) |
| HS Completer | 0.959 | 0.898 | 0.962 | 0.978 | 0.933 | 0.970 |
| | (0.197) | (0.303) | (0.191) | (0.146) | (0.250) | (0.170) |
| Dropout | 0.036 | 0.089 | 0.034 | 0.021 | 0.063 | 0.028 |
| | (0.187) | (0.285) | (0.180) | (0.143) | (0.244) | (0.165) |
| *Panel E: Postsecondary Enrollment* | | | | | | |
| College Enrollment (16 mo.) | 0.599 | 0.369 | 0.515 | 0.654 | 0.346 | 0.500 |
| | (0.490) | (0.483) | (0.500) | (0.476) | (0.476) | (0.500) |
| 4-Year College Enrollment | 0.418 | 0.168 | 0.282 | 0.449 | 0.122 | 0.227 |
| | (0.493) | (0.374) | (0.450) | (0.497) | (0.328) | (0.419) |
| 2-Year College Enrollment | 0.330 | 0.290 | 0.360 | 0.356 | 0.299 | 0.385 |
| | (0.470) | (0.454) | (0.480) | (0.479) | (0.458) | (0.487) |

*Notes:* Cells report the mean, with standard deviation in parentheses; demographic rows in Panel A are percentages except where noted. *Full Sample* includes all students in the analysis sample. *Failed (<400)* restricts to students who scored below the passing cutoff on their first attempt. *BW [380, 420]* restricts to students within ±20 points of the cutoff, the preferred estimation window. Algebra I sample: TC120 (2009 Standards), graduation cohorts 2016–2019. English sample: TC109, graduation cohorts 2014–2019. Race/ethnicity percentages may not sum to 100 due to other categories not shown. Panel B outcomes are not available for English because they measure downstream performance on math-specific SOL exams. Attendance Year 2 is likewise Algebra-I-only: English students take the exam in 11th grade, so a second post-exam attendance year is not consistently observed before graduation. *4-Year College Enrollment* is enrollment at a 4-year institution within 2 years of HS graduation.

Table 2. Effect of Failing Algebra I on Downstream Math Progression: Linear RD

| | Geometry SOL | | | | Algebra II SOL | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) Took | (2) Score (cond.) | (3) Pass (cond.) | (4) Pass (unc.) | (5) Took | (6) Score (cond.) | (7) Pass (cond.) | (8) Pass (unc.) |
| **Panel A: Bandwidth Sensitivity (Full Sample)** | | | | | | | | |
| [390, 410] | 0.016*** | 4.014*** | 0.052*** | 0.054*** | 0.045*** | 3.029*** | 0.029*** | 0.046*** |
| | (0.006) | (0.736) | (0.010) | (0.009) | (0.009) | (1.121) | (0.011) | (0.009) |
| [385, 415] | -0.001 | 2.911*** | 0.036*** | 0.027*** | 0.016** | 3.537*** | 0.029*** | 0.026*** |
| | (0.005) | (0.551) | (0.008) | (0.007) | (0.007) | (0.893) | (0.009) | (0.007) |
| **[380, 420]** | **0.008*** | **2.572***** | **0.028***** | **0.026***** | **0.017***** | **2.341***** | **0.021***** | **0.021***** |
| | **(0.004)** | **(0.468)** | **(0.007)** | **(0.006)** | **(0.006)** | **(0.759)** | **(0.008)** | **(0.006)** |
| [375, 425] | 0.007** | 2.561*** | 0.023*** | 0.021*** | 0.010** | 2.425*** | 0.021*** | 0.017*** |
| | (0.004) | (0.404) | (0.006) | (0.005) | (0.005) | (0.684) | (0.007) | (0.005) |
| Dep. var. mean | 0.876 | 406.443 | 0.570 | 0.487 | 0.614 | 416.843 | 0.660 | 0.373 |
| Observations | 118,335 | 101,062 | 101,062 | 118,335 | 118,335 | 66,832 | 66,832 | 118,335 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | | | |
| **Black & Hispanic** | | | | | | | | |
| | 0.008 | 2.318*** | 0.025** | 0.023** | 0.019** | 2.597** | 0.024** | 0.022*** |
| | (0.006) | (0.676) | (0.010) | (0.009) | (0.008) | (1.129) | (0.012) | (0.008) |
| Dep. var. mean | 0.882 | 397.973 | 0.469 | 0.400 | 0.588 | 410.510 | 0.610 | 0.324 |
| Observations | 50,306 | 42,965 | 42,965 | 50,306 | 50,306 | 26,743 | 26,743 | 50,306 |
| **Economically Disadvantaged** | | | | | | | | |
| | 0.005 | 2.612*** | 0.032*** | 0.025*** | 0.024*** | 2.317* | 0.009 | 0.014* |
| | (0.006) | (0.672) | (0.010) | (0.008) | (0.008) | (1.216) | (0.013) | (0.008) |
| Dep. var. mean | 0.879 | 398.458 | 0.474 | 0.398 | 0.516 | 410.982 | 0.614 | 0.282 |
| Observations | 53,968 | 45,346 | 45,346 | 53,968 | 53,968 | 24,800 | 24,800 | 53,968 |
| **Male** | | | | | | | | |
| | 0.008 | 2.866*** | 0.038*** | 0.035*** | 0.016** | 3.578*** | 0.035*** | 0.028*** |
| | (0.006) | (0.679) | (0.009) | (0.008) | (0.008) | (1.113) | (0.011) | (0.008) |
| Dep. var. mean | 0.874 | 407.541 | 0.583 | 0.495 | 0.580 | 416.852 | 0.660 | 0.350 |
| Observations | 60,693 | 51,470 | 51,470 | 60,693 | 60,693 | 32,157 | 32,157 | 60,693 |
| Cohort FEs | Yes | | | | | | | |
| Linear RD poly | Yes | | | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ (effect of failing) from the linear RD at cutoff 400, full sample, one row per bandwidth ($h \in \{10, 15, 20, 25\}$); preferred bandwidth [380, 420] is bolded. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. For each subject: "Took"= 1 if the student ever enrolled in that course after the Algebra I SOL (unconditional); "Score (cond.)"= scaled score (0–600) on the first downstream SOL attempt, defined only among takers (conditional, selection-prone); "Pass (cond.)"= 1 if the first downstream SOL attempt passed, defined only among takers (conditional, selection-prone because failing raises take-up); "Pass (unc.)"= 1 if the student sat *and* passed that SOL, coding non-takers as 0 (full sample, selection-free). Cohort fixed effects (cohort 2016 base) and school-clustered SEs. Sample: TC120 (2009 Standards), graduation cohorts 2016–2019. Panel A dep. var. mean and observations are from the preferred bandwidth. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3. Effect of Failing the SOL Exam on High School Completion and Postsecondary Enrollment: Linear RD

| | High School Completion | | | | | | Postsecondary Enrollment | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Algebra I (TC120) | | | English (TC109) | | | Algebra I (TC120) | | | English (TC109) | | |
| | (1) HS Grad. | (2) HS Compl. | (3) Dropout | (4) HS Grad. | (5) HS Compl. | (6) Dropout | (7) Coll. 16mo | (8) 4-yr Enroll. | (9) 2-yr Enroll. | (10) Coll. 16mo | (11) 4-yr Enroll. | (12) 2-yr Enroll. |
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | | | | | | | |
| [390, 410] | 0.002 | -0.002 | 0.002 | -0.005 | -0.003 | 0.003 | 0.027*** | 0.013 | 0.025*** | 0.008 | 0.004 | 0.008 |
| | (0.004) | (0.003) | (0.003) | (0.003) | (0.003) | (0.003) | (0.009) | (0.008) | (0.008) | (0.008) | (0.007) | (0.008) |
| [385, 415] | -0.001 | -0.003 | 0.003 | -0.001 | 0.000 | -0.000 | 0.012* | 0.006 | 0.010 | 0.009 | 0.003 | 0.008 |
| | (0.004) | (0.003) | (0.003) | (0.003) | (0.002) | (0.002) | (0.007) | (0.007) | (0.006) | (0.007) | (0.005) | (0.007) |
| **[380, 420]** | **0.002** | **-0.002** | **0.002** | **0.000** | **0.001** | **-0.000** | **0.010*** | **0.008** | **0.007** | **0.004** | **0.004** | **-0.000** |
| | **(0.003)** | **(0.003)** | **(0.002)** | **(0.002)** | **(0.002)** | **(0.002)** | **(0.006)** | **(0.006)** | **(0.005)** | **(0.006)** | **(0.005)** | **(0.006)** |
| [375, 425] | 0.003 | -0.000 | 0.001 | 0.001 | 0.001 | -0.001 | 0.010* | 0.007 | 0.005 | 0.005 | 0.003 | 0.001 |
| | (0.003) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.006) | (0.005) | (0.005) | (0.005) | (0.004) | (0.005) |
| Dep. var. mean | 0.943 | 0.962 | 0.034 | 0.955 | 0.970 | 0.028 | 0.515 | 0.282 | 0.360 | 0.500 | 0.227 | 0.385 |
| Observations | 108,041 | 108,041 | 118,335 | 115,825 | 115,825 | 119,545 | 118,335 | 118,335 | 118,335 | 119,536 | 119,536 | 119,536 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | | | | | | | |
| **Black & Hispanic** | | | | | | | | | | | | |
| | -0.001 | -0.003 | 0.004 | -0.003 | -0.001 | 0.001 | 0.003 | -0.001 | 0.008 | -0.006 | -0.004 | -0.003 |
| | (0.005) | (0.004) | (0.004) | (0.003) | (0.003) | (0.003) | (0.008) | (0.008) | (0.008) | (0.008) | (0.007) | (0.008) |
| Dep. var. mean | 0.937 | 0.955 | 0.040 | 0.956 | 0.971 | 0.028 | 0.497 | 0.270 | 0.347 | 0.500 | 0.235 | 0.374 |
| Observations | 46,130 | 46,130 | 50,306 | 57,551 | 57,551 | 59,290 | 50,306 | 50,306 | 50,306 | 59,282 | 59,282 | 59,282 |
| **Economically Disadvantaged** | | | | | | | | | | | | |
| | -0.001 | -0.007 | 0.006 | -0.003 | -0.002 | 0.001 | -0.001 | -0.003 | -0.004 | -0.003 | -0.006 | 0.008 |
| | (0.005) | (0.004) | (0.004) | (0.004) | (0.003) | (0.003) | (0.009) | (0.007) | (0.008) | (0.008) | (0.006) | (0.008) |
| Dep. var. mean | 0.914 | 0.940 | 0.054 | 0.942 | 0.960 | 0.038 | 0.411 | 0.183 | 0.325 | 0.429 | 0.169 | 0.350 |
| Observations | 49,273 | 49,273 | 53,968 | 59,416 | 59,416 | 61,229 | 53,968 | 53,968 | 53,968 | 61,222 | 61,222 | 61,222 |
| **Male** | | | | | | | | | | | | |
| | 0.001 | -0.002 | 0.002 | -0.000 | -0.000 | 0.000 | 0.005 | 0.011 | -0.002 | 0.002 | 0.005 | -0.007 |
| | (0.004) | (0.004) | (0.003) | (0.004) | (0.003) | (0.003) | (0.008) | (0.007) | (0.008) | (0.008) | (0.006) | (0.008) |
| Dep. var. mean | 0.934 | 0.956 | 0.039 | 0.946 | 0.964 | 0.034 | 0.454 | 0.236 | 0.325 | 0.437 | 0.193 | 0.338 |
| Observations | 55,490 | 55,490 | 60,693 | 60,641 | 60,641 | 62,580 | 60,693 | 60,693 | 60,693 | 62,571 | 62,571 | 62,571 |
| Cohort FEs | | | | | | | Yes | | | | | |
| Linear RD poly | | | | | | | Yes | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $\widetilde{S}_i$ is the first-attempt score centered at 400 and the interaction allows the linear slope to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 400$. Columns 1–6 are High School Completion outcomes (Algebra I then English); columns 7–12 are Postsecondary Enrollment outcomes (Algebra I then English). Cohort fixed effects included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; English — TC109, cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

## Table 4. Effect of Failing the SOL Exam on Mechanisms: Linear RD

| | Algebra I (TC120) | | | | | | | English (TC109) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| | Att. Yr 1 | Att. Yr 2 | CTE Fam. After | CTE Comp. After | ACT/SAT | AP | Disab. After | Att. Yr 1 | CTE Fam. After | CTE Comp. After | ACT/SAT | AP | Disab. After |
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | | | | | | | | |
| [390, 410] | -0.001 | 0.001 | 0.015*** | -0.012 | 0.014 | 0.016* | 0.003 | 0.001 | -0.000 | 0.004 | -0.016 | 0.005 | 0.004** |
| | (0.001) | (0.002) | (0.006) | (0.009) | (0.010) | (0.009) | (0.003) | (0.002) | (0.007) | (0.009) | (0.010) | (0.007) | (0.002) |
| [385, 415] | -0.001 | 0.001 | 0.003 | -0.009 | 0.009 | 0.011 | -0.002 | 0.001 | 0.010 | 0.013* | -0.005 | 0.005 | 0.003** |
| | (0.001) | (0.001) | (0.005) | (0.007) | (0.008) | (0.007) | (0.002) | (0.001) | (0.006) | (0.008) | (0.008) | (0.005) | (0.001) |
| **[380, 420]** | **-0.001** | **0.001** | **0.007*** | **-0.006** | **0.003** | **0.011*** | **-0.002** | **0.000** | **0.005** | **0.011*** | **-0.002** | **0.001** | **0.002*** |
| | **(0.001)** | **(0.001)** | **(0.004)** | **(0.006)** | **(0.007)** | **(0.006)** | **(0.002)** | **(0.001)** | **(0.005)** | **(0.006)** | **(0.007)** | **(0.004)** | **(0.001)** |
| [375, 425] | -0.000 | 0.002* | 0.005 | -0.004 | -0.001 | 0.012** | -0.000 | -0.000 | 0.004 | 0.006 | -0.005 | 0.000 | 0.001 |
| | (0.001) | (0.001) | (0.004) | (0.005) | (0.007) | (0.006) | (0.002) | (0.001) | (0.005) | (0.006) | (0.006) | (0.004) | (0.001) |
| Dep. var. mean | 0.942 | 0.933 | 0.863 | 0.467 | 0.581 | 0.252 | 0.032 | 0.918 | 0.675 | 0.521 | 0.530 | 0.129 | 0.009 |
| Observations | 114,239 | 107,584 | 118,335 | 114,379 | 85,302 | 85,281 | 103,130 | 107,026 | 119,545 | 105,111 | 82,747 | 82,732 | 99,110 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | | | | | | | | |
| **Black & Hispanic** | | | | | | | | | | | | | |
| | -0.001 | 0.000 | 0.012** | -0.000 | -0.004 | -0.003 | -0.002 | -0.002 | 0.003 | 0.014 | -0.016* | -0.008 | 0.003** |
| | (0.001) | (0.002) | (0.006) | (0.008) | (0.011) | (0.008) | (0.003) | (0.002) | (0.008) | (0.009) | (0.009) | (0.006) | (0.001) |
| Dep. var. mean | 0.942 | 0.931 | 0.869 | 0.457 | 0.562 | 0.202 | 0.024 | 0.917 | 0.663 | 0.501 | 0.535 | 0.119 | 0.007 |
| Observations | 48,392 | 45,424 | 50,306 | 48,476 | 37,638 | 37,631 | 44,266 | 52,501 | 59,290 | 51,655 | 43,645 | 43,639 | 50,774 |
| **Economically Disadvantaged** | | | | | | | | | | | | | |
| | -0.003* | -0.000 | 0.005 | -0.009 | -0.013 | 0.004 | -0.001 | 0.001 | 0.003 | 0.007 | -0.015 | -0.003 | 0.001 |
| | (0.001) | (0.002) | (0.006) | (0.008) | (0.011) | (0.008) | (0.003) | (0.002) | (0.008) | (0.008) | (0.010) | (0.007) | (0.002) |
| Dep. var. mean | 0.930 | 0.918 | 0.865 | 0.469 | 0.482 | 0.169 | 0.024 | 0.903 | 0.663 | 0.523 | 0.461 | 0.110 | 0.008 |
| Observations | 52,268 | 48,993 | 53,968 | 52,150 | 34,708 | 34,703 | 46,773 | 53,682 | 61,229 | 52,343 | 39,521 | 39,512 | 51,870 |
| **Male** | | | | | | | | | | | | | |
| | -0.000 | 0.001 | 0.004 | -0.012 | 0.010 | 0.013 | -0.003 | -0.002 | 0.002 | 0.007 | 0.010 | -0.002 | 0.001 |
| | (0.001) | (0.002) | (0.005) | (0.008) | (0.011) | (0.008) | (0.003) | (0.002) | (0.007) | (0.008) | (0.009) | (0.005) | (0.002) |
| Dep. var. mean | 0.946 | 0.937 | 0.877 | 0.501 | 0.537 | 0.212 | 0.032 | 0.921 | 0.695 | 0.533 | 0.491 | 0.104 | 0.009 |
| Observations | 58,534 | 55,097 | 60,693 | 58,652 | 41,937 | 41,924 | 51,088 | 56,154 | 62,580 | 55,196 | 41,407 | 41,396 | 49,749 |
| Cohort FEs | | | | | | | Yes | | | | | | |
| Linear RD poly | | | | | | | Yes | | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $\widetilde{S}_i$ is the first-attempt score centered at 400 and the interaction allows the linear slope to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 400$. Columns 1–7 are Algebra I; columns 8–13 are English (Reading), which lacks an Attendance Year 2 outcome. Cohort fixed effects included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. CTE Fam. After = 1 if the student enrolled in any CTE career family in any year strictly after the first SOL attempt. CTE Comp. After = 1 if the student attained CTE competency in any year strictly after the first SOL attempt. ACT/SAT = 1 if the student took the ACT or SAT in any year strictly after the first SOL attempt. AP = 1 if the student took any AP exam in any year strictly after the first SOL attempt. Disab. After = 1 if the student (not disabled at time of test) received any new disability classification in any school year after the first SOL attempt. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; English — TC109, cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5. Effect of Scoring Below 375 on the SOL Exam on High School Completion and Postsecondary Enrollment: Linear RD

| | **High School Completion** | | | | | | **Postsecondary Enrollment** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Algebra I (TC120) | | | English (TC109) | | | Algebra I (TC120) | | | English (TC109) | | |
| | (1) HS Grad. | (2) HS Compl. | (3) Dropout | (4) HS Grad. | (5) HS Compl. | (6) Dropout | (7) Coll. 16mo | (8) 4-yr Enroll. | (9) 2-yr Enroll. | (10) Coll. 16mo | (11) 4-yr Enroll. | (12) 2-yr Enroll. |
| **Panel A: Bandwidth Sensitivity (cutoff = 375)** | | | | | | | | | | | | |
| [365, 385] | -0.012 | -0.003 | 0.001 | -0.002 | -0.003 | 0.005 | 0.015 | 0.007 | 0.008 | -0.011 | -0.010 | 0.008 |
| | (0.008) | (0.007) | (0.006) | (0.010) | (0.007) | (0.007) | (0.010) | (0.008) | (0.010) | (0.012) | (0.008) | (0.012) |
| [360, 390] | -0.007 | -0.001 | -0.001 | -0.006 | -0.007 | 0.007 | 0.009 | 0.008 | 0.005 | -0.013 | -0.009 | -0.003 |
| | (0.007) | (0.006) | (0.005) | (0.008) | (0.005) | (0.005) | (0.008) | (0.007) | (0.008) | (0.009) | (0.006) | (0.009) |
| **[355, 395]** | **-0.015*** | **-0.007** | **0.005** | **-0.003** | **-0.004** | **0.003** | **0.003** | **0.004** | **0.003** | **-0.012** | **-0.008** | **-0.006** |
| | **(0.006)** | **(0.005)** | **(0.004)** | **(0.006)** | **(0.005)** | **(0.004)** | **(0.007)** | **(0.006)** | **(0.007)** | **(0.008)** | **(0.005)** | **(0.008)** |
| [351, 399] | -0.015*** | -0.009** | 0.007* | 0.001 | -0.003 | 0.003 | 0.000 | 0.001 | 0.002 | -0.015** | -0.007 | -0.010 |
| | (0.005) | (0.004) | (0.004) | (0.006) | (0.004) | (0.004) | (0.007) | (0.005) | (0.006) | (0.007) | (0.005) | (0.007) |
| Dep. var. mean | 0.858 | 0.911 | 0.078 | 0.891 | 0.944 | 0.053 | 0.381 | 0.172 | 0.299 | 0.369 | 0.130 | 0.318 |
| Observations | 69,761 | 69,761 | 78,083 | 59,572 | 59,572 | 62,054 | 78,076 | 78,076 | 78,076 | 62,041 | 62,041 | 62,041 |
| **Panel B: Subgroup Heterogeneity (BW [355, 395])** | | | | | | | | | | | | |
| **Black & Hispanic** | | | | | | | | | | | | |
| | -0.008 | 0.002 | -0.002 | -0.004 | -0.004 | 0.003 | -0.001 | -0.007 | 0.005 | -0.012 | -0.018** | 0.000 |
| | (0.008) | (0.007) | (0.006) | (0.008) | (0.006) | (0.005) | (0.011) | (0.008) | (0.010) | (0.010) | (0.007) | (0.010) |
| Dep. var. mean | 0.843 | 0.897 | 0.092 | 0.892 | 0.945 | 0.052 | 0.374 | 0.172 | 0.288 | 0.372 | 0.142 | 0.311 |
| Observations | 34,136 | 34,136 | 37,973 | 34,242 | 34,242 | 35,613 | 37,967 | 37,967 | 37,967 | 35,601 | 35,601 | 35,601 |
| **Economically Disadvantaged** | | | | | | | | | | | | |
| | -0.013 | -0.005 | 0.004 | -0.006 | -0.003 | 0.002 | -0.002 | 0.002 | -0.002 | -0.014 | -0.016** | 0.001 |
| | (0.009) | (0.007) | (0.007) | (0.008) | (0.006) | (0.006) | (0.009) | (0.006) | (0.009) | (0.009) | (0.006) | (0.009) |
| Dep. var. mean | 0.808 | 0.876 | 0.109 | 0.875 | 0.932 | 0.065 | 0.300 | 0.113 | 0.255 | 0.325 | 0.106 | 0.287 |
| Observations | 37,688 | 37,688 | 42,097 | 36,527 | 36,527 | 37,941 | 42,090 | 42,090 | 42,090 | 37,933 | 37,933 | 37,933 |
| **Male** | | | | | | | | | | | | |
| | -0.030*** | -0.014** | 0.012** | -0.006 | -0.004 | 0.003 | 0.003 | 0.011 | -0.003 | -0.009 | -0.003 | -0.008 |
| | (0.008) | (0.006) | (0.006) | (0.008) | (0.006) | (0.006) | (0.009) | (0.007) | (0.009) | (0.011) | (0.006) | (0.010) |
| Dep. var. mean | 0.838 | 0.897 | 0.090 | 0.877 | 0.936 | 0.060 | 0.325 | 0.138 | 0.258 | 0.319 | 0.111 | 0.273 |
| Observations | 38,791 | 38,791 | 43,310 | 32,337 | 32,337 | 33,688 | 43,303 | 43,303 | 43,303 | 33,675 | 33,675 | 33,675 |
| Cohort FEs | | | | | | | Yes | | | | | |
| Linear RD poly | | | | | | | Yes | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 375, where $\widetilde{S}_i$ is the first-attempt score centered at 375 and the interaction allows the linear slope to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 375$. Columns 1–6 are High School Completion outcomes (Algebra I then English); columns 7–12 are Postsecondary Enrollment outcomes (Algebra I then English). Cohort fixed effects included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [355, 395]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; scores $\leq 399$; English — TC109, cohorts 2014–2019; scores $\leq 399$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

## References


**Applequist, K.** 2008. "No Child Left Behind Act (2001)." 1444–1446. 10.1002/9780470373699.speced1479.

**Bergman, Peter.** 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Journal of Political Economy* 129 (1): 286–322. 10.1086/711410.

**Clark, Damon, and Francisco Martorell.** 2014. "The Signaling Value of a High School Diploma." *Journal of Political Economy* 122 (2): 282–318. 10.1086/675238.

**Cortes, Kalena, Joshua Goodman, and Takako Nomi.** 2014. "Intensive Math Instruction and Educational Attainment: Long-Run Impacts of Double-Dose Algebra." Working Paper20211, National Bureau of Economic Research. 10.3386/w20211.

**Dee, Thomas S., and Brian Jacob.** 2011. "The Impact of No Child Left Behind on Student Achievement." *Journal of Policy Analysis and Management* 30 (3): 418–446, http://www.jstor.org/stable/23018959.

**FairTest.** 2017. "Graduation Test Update: States That Recently Eliminated or Scaled Back High School Exit Exams." January, https://fairtest.org/graduation-test-update-states-recently-eliminated/.

**Figlio, David, and Umut Ozek.** 2024. "The Unintended Consequences of Test-Based Remediation." *American Economic Journal: Applied Economics* 16 (1): 60–89. 10.1257/app.20210037.

**Hanushek, Eric A., and Margaret E. Raymond.** 2005. "Does school accountability lead to improved student performance?" *Journal of Policy Analysis and Management* 24 (2): 297–327. https://doi.org/10.1002/pam.20091.

**Jacob, Brian A.** 2005. "Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools." *Journal of public Economics* 89 (5-6): 761–796.

**Jacob, Brian A., and Lars Lefgren.** 2009. "The Effect of Grade Retention on High School Completion." *American Economic Journal: Applied Economics* 1 (3): 33–58. 10.1257/app.1.3.33.

**Lee, David S., and Thomas Lemieux.** 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355. 10.1257/jel.48.2.281.

**Manacorda, Marco.** 2012. "THE COST OF GRADE RETENTION." *The Review of Economics and Statistics* 94 (2): 596–606, http://www.jstor.org/stable/23262090.

**McGuinn, Patrick.** 2016. "From no child left behind to the every student succeeds act: Federalism and the education legacy of the Obama administration." *Publius: The Journal of Federalism* 46 (3): 392–415.

**McKenna, Marla.** 2018. "State changes rules for locally-verified credits, opens door for retroactive diplomas." https://starexponent.com/news/state-changes-rules-for-locally-verified-credits-opens-door-for/article_ef6c034f-72a9-50a1-8d73-130c2b47c75c.html.

**Neal, Derek, and Diane Whitmore Schanzenbach.** 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92 (2): 263—-283. 10.1162/rest.2010.12318.

**Özek, Umut.** 2025. "The Effects of High School Remediation on Long-Run Educational Attainment." *Journal of Human Resources.*

**Papay, John P., Richard J. Murnane, and John B. Willett.** 2010. "The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence From

Massachusetts." *Educational Evaluation and Policy Analysis* 32 (1): 5–23. 10.3102/0162373709352530.

**Reardon, Sean F., Nicole Arshan, Allison Atteberry, and Michal Kurlaender.** 2010. "Effects of Failing a High School Exit Exam on Course Taking, Achievement, Persistence, and Graduation." *Educational Evaluation and Policy Analysis* 32 (4): 498–520. 10.3102/0162373710382655.

**Schwerdt, Guido, Martin R West, and Marcus A Winters.** 2017. "The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida." *Journal of Public Economics* 152 154–169.

**U.S. Department of Education.** 2007. "Virginia Assessment Letter." December, https://www.ed.gov/laws-and-policy/laws-preschool-grade-12-education/preschool-grade-12-policy-documents/virginia-assessment-letter, Office of Elementary and Secondary Education. Accessed: 2026-06-20.

**Virginia Department of Education.** 2021. "Guidance Document Governing Certain Provisions of the Regulations Establishing Standards for Accrediting Public Schools in Virginia (8VAC20-131)." technical report, Virginia Department of Education, https://www.doe.virginia.gov/home/showpublisheddocument/52591/638884565887170000, Revised March 2021. Accessed: 2026-05-02.

**Virginia Department of Education.** 2024. "Standards of Learning Assessments: Retesting." Virginia Administrative Code, 8 VAC 20-132-40, https://law.lis.virginia.gov/admincode/title8/agency20/chapter132/section40/, Accessed April 21, 2026.

**Virginia Department of Education.** 2025. "Revised State Template for the Consolidated State Plan: The Elementary and Secondary Education Act of 1965, as Amended by the Every Student Succeeds Act." technical report, Virginia Department of Education, https://www.doe.virginia.gov/home/showpublisheddocument/52591/638884565887170000, Original application approved July 27, 2017; Amendment 8 submitted May 22, 2025. Accessed: 2026-05-02.

**Virginia Department of Education.** n.d.a. "SOL Practice Items (All Subjects)." https://www.doe.virginia.gov/teaching-learning-assessment/student-assessment/sol-practice-items-all-subjects, Accessed: 2026-05-02.

**Virginia Department of Education.** n.d.b. "SOL Test Scoring Performance Reports." https://www.doe.virginia.gov/teaching-learning-assessment/student-assessment/virginia-sol-assessment-program/sol-test-scoring-performance-reports, Accessed: 2026-05-02.

**Virginia Department of Education.** n.d.c. "Virginia SOL Assessment Program." https://www.doe.virginia.gov/teaching-learning-assessment/student-assessment/virginia-sol-assessment-program, Accessed: 2026-05-02.

**Yecke, Cheri Pierson.** 1999. "SOLs: Fact and Fiction." Keynote Address presented at the Virginia State PTA Leadership Training Convention. Virginia Department of Education. http://www.knowledge.state.va.us/main/news/solsfactfiction.htm, July, Archived by the Wayback Machine, October 25, 2008. https://web.archive.org/web/20081025093301/http://www.knowledge.state.va.us/main/news/solsfactfiction.htm.

**Zabala, Dalia, Angela Minnici, Jennifer McMurrer, and Liza Briggs.** 2008. *State High School Exit Exams: A Move toward End-of-Course Exams.* https://eric.ed.gov/?id=ED504468, ERIC Number: ED504468.

# A Appendix Tables

Table A1. SOL Retaking and Passing Behavior by Test Code, Federal Graduation Cohorts 2010–2019

| | Algebra I | | | English | | |
|---|---|---|---|---|---|---|
| | Overall (1) | TC 20 (2) | TC 120 (3) | Overall (4) | TC 30 (5) | TC 109 (6) |
| Students ($N$) | 932,413 | 491,015 | 418,103 | 876,308 | 347,183 | 529,059 |
| Passed on first attempt (%) | 80.2 | 92.2 | 67.7 | 85.8 | 90.8 | 82.5 |
| *Among first-attempt failers* | | | | | | |
| *Panel A: Retaking Behavior* | | | | | | |
| Retook same school year (%) | 62.2 | 58.4 | 63.9 | 67.4 | 60.1 | 69.8 |
| First retook in later school year (%) | 27.3 | 23.3 | 26.5 | 22.4 | 26.6 | 20.6 |
| Never retook (%) | 10.4 | 18.2 | 9.6 | 10.3 | 13.3 | 9.6 |
| *Panel B: Passing Outcomes* | | | | | | |
| Passed by end of first school year (%) | 28.7 | 29.9 | 29.2 | 33.4 | 38.2 | 31.7 |
| Passed in later school year (%) | 32.8 | 32.8 | 32.3 | 22.1 | 33.7 | 17.8 |
| Never passed (%) | 38.4 | 37.3 | 38.5 | 44.5 | 28.1 | 50.5 |

*Notes:* Sample includes all students with at least one SOL attempt in federal graduation cohorts 2010–2019. Columns (1) and (4) include all test codes for Algebra I and English, respectively. TC 20 and TC 120 correspond to the Algebra I end-of-course SOL; TC 30 and TC 109 correspond to the English end-of-course SOL. Row 2 reports the share who passed on their first attempt (score $\geq$ 400). All remaining rows condition on failing the first attempt. Panel A rows sum to 100%: *retook same school year* = any retake attempt in the same school year as the first attempt; *first retook in later school year* = no same-year retake but at least one attempt in a subsequent school year; *never retook* = no additional attempts ever recorded. Panel B rows sum to 100%: *passed by end of first school year* = achieved a passing score ($\geq$ 400) via a same-year retake; *passed in later school year* = no same-year pass but a passing score recorded in a subsequent school year; *never passed* = no passing score ever recorded.

Table A2. Covariate Smoothness at the Algebra I Passing Cutoff: Linear RD (TC120, Cohorts 2016–2019)

| | [390, 410] | [385, 415] | **[380, 420]** | [375, 425] |
|---|---|---|---|---|
| Female | -0.008 | -0.006 | -0.006 | -0.008 |
| | (0.008) | (0.007) | (0.006) | (0.006) |
| Black | -0.015* | -0.008 | -0.001 | -0.006 |
| | (0.009) | (0.007) | (0.006) | (0.005) |
| White | 0.000 | 0.002 | -0.004 | -0.002 |
| | (0.009) | (0.007) | (0.006) | (0.005) |
| Hispanic | 0.007 | 0.004 | 0.003 | 0.005 |
| | (0.006) | (0.005) | (0.004) | (0.004) |
| Asian | 0.005* | 0.003 | 0.002 | 0.002 |
| | (0.003) | (0.002) | (0.002) | (0.002) |
| Econ. Disadvantaged | -0.017* | -0.004 | -0.003 | -0.004 |
| | (0.009) | (0.006) | (0.005) | (0.005) |
| Joint test $p$-value | 0.057 | 0.605 | 0.679 | 0.246 |
| Observations | 61,464 | 88,180 | 118,335 | 143,520 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 400, and $D_i = 1$ if first Algebra I score $< 400$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2016 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Sample: TC120, graduation cohorts 2016–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A3. Covariate Smoothness at the English (Reading) Passing Cutoff: Linear RD (TC109, Cohorts 2014–2019)

| | [390, 410] | [385, 415] | **[380, 420]** | [375, 425] |
|---|---|---|---|---|
| Female | -0.005 | 0.003 | 0.004 | 0.006 |
| | (0.009) | (0.007) | (0.006) | (0.005) |
| Black | -0.011 | -0.008 | -0.004 | -0.002 |
| | (0.009) | (0.008) | (0.006) | (0.006) |
| White | 0.013 | 0.010 | 0.003 | 0.005 |
| | (0.009) | (0.008) | (0.006) | (0.006) |
| Hispanic | -0.004 | -0.003 | 0.001 | -0.003 |
| | (0.006) | (0.005) | (0.004) | (0.004) |
| Asian | 0.004 | 0.004 | 0.003 | 0.003 |
| | (0.004) | (0.003) | (0.003) | (0.002) |
| Econ. Disadvantaged | -0.010 | 0.001 | 0.012* | 0.007 |
| | (0.009) | (0.007) | (0.006) | (0.005) |
| Joint test $p$-value | 0.523 | 0.426 | 0.192 | 0.178 |
| Observations | 58,171 | 88,737 | 119,545 | 148,767 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 400, and $D_i = 1$ if first English (Reading) score $< 400$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2014 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Sample: TC109, graduation cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A4. Covariate Smoothness at the Algebra I Expedited-Retake Cutoff: Linear RD (TC120, Cohorts 2016–2019)

| | [365, 385] | [360, 390] | **[355, 395]** | [351, 399] |
|---|---|---|---|---|
| Female | -0.005 | -0.004 | -0.002 | -0.003 |
| | (0.011) | (0.009) | (0.008) | (0.007) |
| Black | -0.001 | 0.005 | 0.001 | 0.002 |
| | (0.010) | (0.009) | (0.007) | (0.007) |
| White | -0.004 | -0.011 | -0.005 | -0.007 |
| | (0.011) | (0.009) | (0.008) | (0.007) |
| Hispanic | 0.003 | 0.006 | 0.003 | 0.004 |
| | (0.008) | (0.006) | (0.006) | (0.005) |
| Asian | 0.003 | 0.005* | 0.004* | 0.004** |
| | (0.003) | (0.003) | (0.002) | (0.002) |
| Econ. Disadvantaged | -0.000 | 0.007 | 0.006 | 0.005 |
| | (0.011) | (0.009) | (0.008) | (0.007) |
| Joint test $p$-value | 0.974 | 0.376 | 0.505 | 0.349 |
| Observations | 39,637 | 57,393 | 78,083 | 93,800 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 375, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 375, and $D_i = 1$ if first Algebra I score $< 375$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2016 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [355, 395]. Sample: TC120, graduation cohorts 2016–2019; sample restricted to scores $\leq 399$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A5. Covariate Smoothness at the English (Reading) Expedited-Retake Cutoff: Linear RD (TC109, Cohorts 2014–2019)

| | [365, 385] | [360, 390] | **[355, 395]** | [351, 399] |
|---|---|---|---|---|
| Female | 0.013 | 0.006 | 0.001 | -0.007 |
| | (0.012) | (0.009) | (0.008) | (0.008) |
| Black | 0.009 | 0.008 | 0.001 | 0.001 |
| | (0.013) | (0.011) | (0.009) | (0.009) |
| White | -0.021* | -0.029*** | -0.014* | -0.014* |
| | (0.012) | (0.009) | (0.008) | (0.008) |
| Hispanic | 0.010 | 0.021** | 0.012* | 0.012* |
| | (0.011) | (0.008) | (0.007) | (0.007) |
| Asian | 0.005 | 0.003 | 0.002 | 0.004 |
| | (0.005) | (0.004) | (0.003) | (0.003) |
| Econ. Disadvantaged | -0.001 | 0.015 | 0.010 | 0.006 |
| | (0.013) | (0.009) | (0.008) | (0.007) |
| Joint test $p$-value | 0.402 | 0.020 | 0.381 | 0.122 |
| Observations | 29,560 | 44,103 | 62,054 | 75,334 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta\widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 375, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 375, and $D_i = 1$ if first English (Reading) score $< 375$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2014 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [355, 395]. Sample: TC109, graduation cohorts 2014–2019; sample restricted to scores $\leq 399$. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

### A.1 375 Remediation Threshold (Linear RD)

Table A6. Effect of Scoring Below 375 on Downstream Math Progression: Linear RD

| | Geometry SOL | | | | Algebra II SOL | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Took | Score (cond.) | Pass (cond.) | Pass (unc.) | Took | Score (cond.) | Pass (cond.) | Pass (unc.) |
| **Panel A: Bandwidth Sensitivity (Full Sample)** | | | | | | | | |
| [365, 385] | -0.002 | 0.699 | 0.003 | 0.002 | -0.018 | -1.617 | 0.015 | 0.000 |
| | (0.009) | (0.887) | (0.012) | (0.010) | (0.011) | (1.907) | (0.020) | (0.008) |
| [360, 390] | 0.000 | 1.011 | 0.011 | 0.004 | -0.012 | -1.967 | 0.004 | -0.006 |
| | (0.007) | (0.716) | (0.010) | (0.008) | (0.009) | (1.514) | (0.016) | (0.006) |
| **[355, 395]** | **-0.006** | **0.925** | **0.013** | **0.003** | **-0.017**** | **0.685** | **0.025*** | **0.001** |
| | **(0.007)** | **(0.608)** | **(0.008)** | **(0.007)** | **(0.007)** | **(1.368)** | **(0.014)** | **(0.005)** |
| [351, 399] | -0.008 | 1.131** | 0.012 | 0.004 | -0.023*** | 0.065 | 0.016 | -0.003 |
| | (0.006) | (0.557) | (0.008) | (0.006) | (0.007) | (1.196) | (0.013) | (0.005) |
| Dep. var. mean | 0.804 | 389.948 | 0.376 | 0.283 | 0.400 | 403.892 | 0.548 | 0.193 |
| Observations | 78,083 | 58,809 | 58,809 | 78,083 | 78,083 | 27,548 | 27,548 | 78,083 |
| **Panel B: Subgroup Heterogeneity (BW [355, 395])** | | | | | | | | |
| **Black & Hispanic** | | | | | | | | |
| | -0.016* | 0.091 | 0.003 | -0.004 | -0.024** | -1.228 | 0.019 | -0.002 |
| | (0.009) | (0.769) | (0.011) | (0.008) | (0.010) | (1.966) | (0.021) | (0.007) |
| Dep. var. mean | 0.806 | 383.426 | 0.298 | 0.223 | 0.379 | 398.390 | 0.501 | 0.164 |
| Observations | 37,973 | 28,320 | 28,320 | 37,973 | 37,973 | 12,455 | 12,455 | 37,973 |
| **Economically Disadvantaged** | | | | | | | | |
| | -0.005 | 0.259 | 0.005 | -0.000 | -0.020** | -1.972 | -0.004 | -0.008 |
| | (0.009) | (0.837) | (0.011) | (0.008) | (0.009) | (2.052) | (0.021) | (0.006) |
| Dep. var. mean | 0.793 | 383.834 | 0.304 | 0.220 | 0.315 | 398.979 | 0.506 | 0.136 |
| Observations | 42,097 | 30,450 | 30,450 | 42,097 | 42,097 | 11,275 | 11,275 | 42,097 |
| **Male** | | | | | | | | |
| | -0.012 | 0.680 | 0.012 | 0.000 | -0.022** | -0.253 | 0.017 | -0.005 |
| | (0.009) | (0.782) | (0.011) | (0.009) | (0.009) | (1.821) | (0.019) | (0.007) |
| Dep. var. mean | 0.795 | 390.176 | 0.383 | 0.283 | 0.365 | 403.419 | 0.547 | 0.174 |
| Observations | 43,310 | 32,089 | 32,089 | 43,310 | 43,310 | 13,798 | 13,798 | 43,310 |
| Cohort FEs | Yes | | | | | | | |
| Linear RD poly | Yes | | | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ (effect of scoring below 375) from the linear RD at cutoff 375, full sample, one row per bandwidth ($h \in \{10, 15, 20, 24\}$); preferred bandwidth [355, 395] is bolded. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. $D_i = 1$ if first Algebra I score $< 375$ (ineligible for the expedited retake); both sides of the cutoff have failed the SOL. For each subject: "Took"$= 1$ if the student ever enrolled in that course after the Algebra I SOL (unconditional); "Score (cond.)"$=$ scaled score (0–600) on the first downstream SOL attempt, defined only among takers (conditional, selection-prone); "Pass (cond.)"$= 1$ if the first downstream SOL attempt passed, defined only among takers (conditional, selection-prone because failing raises take-up); "Pass (unc.)"$= 1$ if the student sat *and* passed that SOL, coding non-takers as 0 (full sample, selection-free). Cohort fixed effects (cohort 2016 base) and school-clustered SEs. Sample: TC120 (2009 Standards), graduation cohorts 2016–2019. Panel A dep. var. mean and observations are from the preferred bandwidth. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A7. Effect of Scoring Below 375 on the SOL Exam on Mechanisms: Linear RD

| | Algebra I (TC120) | | | | | | | English (TC109) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) |
| | Att. Yr 1 | Att. Yr 2 | CTE Fam. After | CTE Comp. After | ACT/SAT | AP | Disab. After | Att. Yr 1 | CTE Fam. After | CTE Comp. After | ACT/SAT | AP | Disab. After |
| **Panel A: Bandwidth Sensitivity (cutoff = 375)** | | | | | | | | | | | | | |
| [365, 385] | 0.000 | -0.001 | -0.009 | -0.002 | -0.007 | -0.006 | 0.009 | -0.000 | -0.020 | -0.019 | -0.003 | 0.000 | -0.002 |
| | (0.002) | (0.003) | (0.008) | (0.011) | (0.014) | (0.009) | (0.006) | (0.003) | (0.013) | (0.014) | (0.016) | (0.008) | (0.004) |
| [360, 390] | -0.001 | -0.001 | -0.011* | -0.003 | -0.013 | -0.003 | 0.005 | -0.003 | -0.016 | -0.007 | 0.004 | -0.003 | -0.001 |
| | (0.002) | (0.002) | (0.006) | (0.009) | (0.012) | (0.008) | (0.005) | (0.003) | (0.010) | (0.011) | (0.013) | (0.006) | (0.003) |
| **[355, 395]** | **-0.002** | **-0.000** | **-0.014**** | **-0.006** | **-0.012** | **-0.002** | **0.005** | **-0.000** | **-0.016*** | **-0.010** | **0.007** | **-0.004** | **-0.000** |
| | **(0.001)** | **(0.002)** | **(0.006)** | **(0.008)** | **(0.010)** | **(0.007)** | **(0.004)** | **(0.002)** | **(0.008)** | **(0.009)** | **(0.011)** | **(0.005)** | **(0.003)** |
| [351, 399] | -0.001 | -0.000 | -0.014*** | -0.009 | -0.013 | 0.001 | 0.004 | -0.001 | -0.006 | -0.006 | 0.004 | -0.004 | -0.002 |
| | (0.001) | (0.002) | (0.005) | (0.007) | (0.009) | (0.006) | (0.004) | (0.002) | (0.007) | (0.009) | (0.010) | (0.004) | (0.003) |
| Dep. var. mean | 0.926 | 0.915 | 0.828 | 0.433 | 0.450 | 0.152 | 0.046 | 0.906 | 0.648 | 0.504 | 0.391 | 0.068 | 0.015 |
| Observations | 74,386 | 68,050 | 78,083 | 74,224 | 48,187 | 48,184 | 62,044 | 53,335 | 62,054 | 51,946 | 36,901 | 36,897 | 44,130 |
| **Panel B: Subgroup Heterogeneity (BW [355, 395])** | | | | | | | | | | | | | |
| **Black & Hispanic** | | | | | | | | | | | | | |
| | -0.003 | -0.001 | -0.020*** | 0.003 | -0.013 | -0.012 | 0.005 | -0.002 | -0.027** | -0.017 | -0.001 | -0.001 | 0.002 |
| | (0.002) | (0.003) | (0.008) | (0.010) | (0.013) | (0.008) | (0.005) | (0.003) | (0.012) | (0.012) | (0.014) | (0.007) | (0.003) |
| Dep. var. mean | 0.923 | 0.911 | 0.828 | 0.413 | 0.446 | 0.121 | 0.037 | 0.905 | 0.632 | 0.473 | 0.409 | 0.065 | 0.012 |
| Observations | 36,001 | 32,723 | 37,973 | 35,953 | 24,767 | 24,767 | 30,503 | 30,214 | 35,613 | 29,504 | 22,819 | 22,817 | 26,690 |
| **Economically Disadvantaged** | | | | | | | | | | | | | |
| | -0.003 | -0.001 | -0.015** | -0.005 | -0.004 | -0.005 | 0.003 | -0.001 | -0.017 | -0.011 | 0.000 | -0.007 | 0.000 |
| | (0.002) | (0.003) | (0.008) | (0.010) | (0.013) | (0.008) | (0.005) | (0.003) | (0.011) | (0.012) | (0.015) | (0.007) | (0.003) |
| Dep. var. mean | 0.912 | 0.898 | 0.820 | 0.418 | 0.374 | 0.100 | 0.040 | 0.893 | 0.630 | 0.494 | 0.351 | 0.064 | 0.013 |
| Observations | 40,251 | 36,561 | 42,097 | 39,963 | 23,231 | 23,231 | 33,063 | 32,004 | 37,941 | 30,941 | 21,455 | 21,452 | 27,933 |
| **Male** | | | | | | | | | | | | | |
| | -0.002 | 0.000 | -0.021*** | -0.007 | -0.014 | -0.002 | 0.006 | -0.002 | -0.004 | -0.004 | 0.019 | -0.007 | -0.004 |
| | (0.002) | (0.003) | (0.008) | (0.010) | (0.012) | (0.008) | (0.005) | (0.003) | (0.012) | (0.012) | (0.016) | (0.006) | (0.004) |
| Dep. var. mean | 0.928 | 0.918 | 0.835 | 0.446 | 0.407 | 0.120 | 0.048 | 0.910 | 0.656 | 0.499 | 0.362 | 0.053 | 0.015 |
| Observations | 41,247 | 37,652 | 43,310 | 41,190 | 25,397 | 25,395 | 32,894 | 28,999 | 33,688 | 28,211 | 19,168 | 19,167 | 22,538 |
| Cohort FEs | | | | | | | Yes | | | | | | |
| Linear RD poly | | | | | | | Yes | | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta \widetilde{S}_i + \delta(\widetilde{S}_i \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 375, where $\widetilde{S}_i$ is the first-attempt score centered at 375 and the interaction allows the linear slope to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 375$. Columns 1–7 are Algebra I; columns 8–13 are English (Reading), which lacks an Attendance Year 2 outcome. Cohort fixed effects included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [355, 395]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. CTE Fam. After = 1 if the student enrolled in any CTE career family in any year strictly after the first SOL attempt. CTE Comp. After = 1 if the student attained CTE competency in any year strictly after the first SOL attempt. ACT/SAT = 1 if the student took the ACT or SAT in any year strictly after the first SOL attempt. AP = 1 if the student took any AP exam in any year strictly after the first SOL attempt. Disab. After = 1 if the student (not disabled at time of test) received any new disability classification in any school year after the first SOL attempt. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; scores $\leq$ 399; English — TC109, cohorts 2014–2019; scores $\leq$ 399. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

## A.2 Quadratic RD Robustness (Cutoff 400)

Table A8. Covariate Smoothness at the Algebra I Passing Cutoff: Quadratic RD (TC120, Cohorts 2016–2019)

| | [390, 410] | [385, 415] | **[380, 420]** | [375, 425] |
|---|---|---|---|---|
| Female | -0.015 | -0.014 | -0.011 | -0.007 |
| | (0.013) | (0.010) | (0.009) | (0.008) |
| Black | -0.009 | -0.019* | -0.020** | -0.006 |
| | (0.014) | (0.012) | (0.009) | (0.008) |
| White | -0.023 | -0.009 | 0.004 | -0.001 |
| | (0.015) | (0.012) | (0.009) | (0.009) |
| Hispanic | 0.017* | 0.012 | 0.008 | 0.003 |
| | (0.010) | (0.008) | (0.006) | (0.005) |
| Asian | 0.006 | 0.008** | 0.005 | 0.004 |
| | (0.004) | (0.003) | (0.003) | (0.003) |
| Econ. Disadvantaged | -0.043*** | -0.033*** | -0.016* | -0.010 |
| | (0.014) | (0.011) | (0.009) | (0.008) |
| Joint test $p$-value | 0.000 | 0.000 | 0.055 | 0.532 |
| Observations | 61,464 | 88,180 | 118,335 | 143,520 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta_1 \widetilde{S}_i + \beta_2 \widetilde{S}_i^2 + \delta_1(\widetilde{S}_i \times D_i) + \delta_2(\widetilde{S}_i^2 \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 400, and $D_i = 1$ if first Algebra I score $< 400$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2016 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Sample: TC120, graduation cohorts 2016–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A9. Covariate Smoothness at the English (Reading) Passing Cutoff: Quadratic RD (TC109, Cohorts 2014–2019)

| | [390, 410] | [385, 415] | **[380, 420]** | [375, 425] |
|---|---|---|---|---|
| Female | 0.013 | -0.003 | 0.001 | 0.002 |
| | (0.013) | (0.011) | (0.009) | (0.008) |
| Black | -0.015 | -0.016 | -0.015 | -0.012 |
| | (0.017) | (0.012) | (0.010) | (0.009) |
| White | 0.032** | 0.018 | 0.023** | 0.016* |
| | (0.016) | (0.012) | (0.010) | (0.009) |
| Hispanic | -0.020* | -0.006 | -0.010 | -0.005 |
| | (0.012) | (0.008) | (0.007) | (0.006) |
| Asian | 0.005 | 0.005 | 0.004 | 0.003 |
| | (0.005) | (0.005) | (0.004) | (0.004) |
| Econ. Disadvantaged | -0.004 | -0.008 | -0.009 | 0.003 |
| | (0.015) | (0.011) | (0.009) | (0.008) |
| Joint test $p$-value | 0.168 | 0.534 | 0.164 | 0.367 |
| Observations | 58,171 | 88,737 | 119,545 | 148,767 |

*Notes:* Each cell reports the OLS estimate of $\tau$ from $x_i = \alpha + \tau D_i + \beta_1 \widetilde{S}_i + \beta_2 \widetilde{S}_i^2 + \delta_1(\widetilde{S}_i \times D_i) + \delta_2(\widetilde{S}_i^2 \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $x_i$ is the covariate in the row label, $\widetilde{S}_i$ is the first-attempt score centered at 400, and $D_i = 1$ if first English (Reading) score $< 400$. Smoothness of observables through the cutoff implies $\tau = 0$. The joint test reports the $p$-value from a seemingly-unrelated (`suest`) Wald test that the discontinuity is zero in all six covariates simultaneously. Cohort fixed effects (cohort 2014 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Sample: TC109, graduation cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A10. Effect of Failing Algebra I on Downstream Math Progression: Quadratic RD (TC120, Cohorts 2016–2019)

| | **Geometry SOL** | | | | **Algebra II SOL** | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | Took | Score (cond.) | Pass (cond.) | Pass (unc.) | Took | Score (cond.) | Pass (cond.) | Pass (unc.) |
| **Full sample** | | | | | | | | |
| | 0.004 | 3.889*** | 0.054*** | 0.048*** | 0.034*** | 3.239*** | 0.030** | 0.041*** |
| | (0.006) | (0.764) | (0.010) | (0.010) | (0.009) | (1.189) | (0.012) | (0.009) |
| Dep. var. mean | 0.876 | 406.443 | 0.570 | 0.487 | 0.614 | 416.843 | 0.660 | 0.373 |
| Observations | 118,335 | 101,062 | 101,062 | 118,335 | 118,335 | 66,832 | 66,832 | 118,335 |
| **Black & Hispanic** | | | | | | | | |
| | 0.002 | 4.129*** | 0.051*** | 0.042*** | 0.033** | 2.160 | 0.028 | 0.037*** |
| | (0.010) | (1.080) | (0.015) | (0.014) | (0.013) | (1.763) | (0.018) | (0.013) |
| Dep. var. mean | 0.882 | 397.973 | 0.469 | 0.400 | 0.588 | 410.510 | 0.610 | 0.324 |
| Observations | 50,306 | 42,965 | 42,965 | 50,306 | 50,306 | 26,743 | 26,743 | 50,306 |
| **Economically Disadvantaged** | | | | | | | | |
| | 0.004 | 4.067*** | 0.053*** | 0.044*** | 0.034*** | 1.955 | 0.017 | 0.025** |
| | (0.009) | (1.091) | (0.014) | (0.013) | (0.013) | (1.882) | (0.019) | (0.012) |
| Dep. var. mean | 0.879 | 398.458 | 0.474 | 0.398 | 0.516 | 410.982 | 0.614 | 0.282 |
| Observations | 53,968 | 45,346 | 45,346 | 53,968 | 53,968 | 24,800 | 24,800 | 53,968 |
| **Male** | | | | | | | | |
| | 0.006 | 4.192*** | 0.069*** | 0.061*** | 0.034*** | 4.200** | 0.035** | 0.039*** |
| | (0.008) | (1.034) | (0.013) | (0.013) | (0.012) | (1.788) | (0.017) | (0.012) |
| Dep. var. mean | 0.874 | 407.541 | 0.583 | 0.495 | 0.580 | 416.852 | 0.660 | 0.350 |
| Observations | 60,693 | 51,470 | 51,470 | 60,693 | 60,693 | 32,157 | 32,157 | 60,693 |
| Cohort FEs | Yes | | | | | | | |
| Quadratic RD poly | Yes | | | | | | | |

*Notes:* OLS estimates of $\tau$ (effect of failing) from the quadratic RD at cutoff 400, preferred bandwidth [380, 420], estimated separately for the full sample and each subgroup. For each subject: "Took"$=1$ if the student ever enrolled in that course after the Algebra I SOL (unconditional); "Score (cond.)"$=$ scaled score (0–600) on the first downstream SOL attempt, defined only among takers (conditional, selection-prone); "Pass (cond.)"$=1$ if the first downstream SOL attempt passed, defined only among takers (conditional, selection-prone because failing raises take-up); "Pass (unc.)"$=1$ if the student sat *and* passed that SOL, coding non-takers as 0 (full sample, selection-free). Cohort fixed effects (cohort 2016 base) and school-clustered SEs. Sample: TC120 (2009 Standards), graduation cohorts 2016–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A11. Effect of Failing the SOL Exam on High School Completion: Quadratic RD

| | Algebra I (TC120) | | | English (TC109) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | HS Grad. | HS Compl. | Dropout | HS Grad. | HS Compl. | Dropout |
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | |
| [390, 410] | 0.006 | 0.000 | 0.000 | -0.001 | 0.003 | -0.004 |
| | (0.006) | (0.005) | (0.004) | (0.005) | (0.005) | (0.004) |
| [385, 415] | 0.006 | -0.001 | 0.001 | -0.005 | -0.001 | 0.001 |
| | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) | (0.003) |
| **[380, 420]** | **0.001** | **-0.002** | **0.002** | **-0.002** | **0.000** | **-0.000** |
| | **(0.004)** | **(0.004)** | **(0.003)** | **(0.004)** | **(0.003)** | **(0.003)** |
| [375, 425] | -0.001 | -0.004 | 0.004 | -0.002 | -0.001 | 0.001 |
| | (0.004) | (0.003) | (0.003) | (0.004) | (0.003) | (0.003) |
| Dep. var. mean | 0.943 | 0.962 | 0.034 | 0.955 | 0.970 | 0.028 |
| Observations | 108,041 | 108,041 | 118,335 | 115,825 | 115,825 | 119,545 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | |
| **Black & Hispanic** | | | | | | |
| | -0.002 | -0.005 | 0.005 | -0.001 | -0.002 | 0.002 |
| | (0.007) | (0.007) | (0.006) | (0.005) | (0.004) | (0.004) |
| Dep. var. mean | 0.937 | 0.955 | 0.040 | 0.956 | 0.971 | 0.028 |
| Observations | 46,130 | 46,130 | 50,306 | 57,551 | 57,551 | 59,290 |
| **Economically Disadvantaged** | | | | | | |
| | 0.002 | -0.003 | 0.003 | -0.008 | -0.007 | 0.007 |
| | (0.008) | (0.007) | (0.006) | (0.006) | (0.005) | (0.005) |
| Dep. var. mean | 0.914 | 0.940 | 0.054 | 0.942 | 0.960 | 0.038 |
| Observations | 49,273 | 49,273 | 53,968 | 59,416 | 59,416 | 61,229 |
| **Male** | | | | | | |
| | 0.001 | -0.001 | 0.001 | 0.003 | -0.000 | 0.000 |
| | (0.006) | (0.006) | (0.005) | (0.006) | (0.005) | (0.005) |
| Dep. var. mean | 0.934 | 0.956 | 0.039 | 0.946 | 0.964 | 0.034 |
| Observations | 55,490 | 55,490 | 60,693 | 60,641 | 60,641 | 62,580 |
| Cohort FEs | Yes | | | | | |
| Quadratic RD poly | Yes | | | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from a regression of the outcome on $D_i$, a quadratic in the centered first-attempt score, their interactions, and cohort fixed effects, estimated within each bandwidth around the cutoff of 400. The interactions allow the quadratic trend to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 400$. Columns 1–3 are Algebra I; columns 4–6 are English (Reading). Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; English — TC109, cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A12. Effect of Failing the SOL Exam on Postsecondary Enrollment: Quadratic RD

| | Algebra I (TC120) | | | English (TC109) | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | Coll. 16mo | 4-yr Enroll. | 2-yr Enroll. | Coll. 16mo | 4-yr Enroll. | 2-yr Enroll. |
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | |
| [390, 410] | 0.034*** | 0.016 | 0.027** | 0.010 | 0.016 | 0.008 |
| | (0.013) | (0.012) | (0.012) | (0.013) | (0.011) | (0.013) |
| [385, 415] | 0.033*** | 0.015 | 0.031*** | 0.010 | 0.011 | 0.008 |
| | (0.010) | (0.010) | (0.010) | (0.010) | (0.008) | (0.010) |
| **[380, 420]** | **0.021**** | **0.007** | **0.021**** | **0.010** | **0.004** | **0.012** |
| | **(0.009)** | **(0.009)** | **(0.009)** | **(0.009)** | **(0.007)** | **(0.009)** |
| [375, 425] | 0.016* | 0.008 | 0.015** | 0.006 | 0.005 | 0.004 |
| | (0.008) | (0.008) | (0.007) | (0.008) | (0.006) | (0.008) |
| Dep. var. mean | 0.515 | 0.282 | 0.360 | 0.500 | 0.227 | 0.385 |
| Observations | 118,335 | 118,335 | 118,335 | 119,536 | 119,536 | 119,536 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | |
| **Black & Hispanic** | | | | | | |
| | 0.006 | -0.010 | 0.027** | 0.014 | 0.006 | 0.013 |
| | (0.014) | (0.012) | (0.013) | (0.012) | (0.010) | (0.012) |
| Dep. var. mean | 0.497 | 0.270 | 0.347 | 0.500 | 0.235 | 0.374 |
| Observations | 50,306 | 50,306 | 50,306 | 59,282 | 59,282 | 59,282 |
| **Economically Disadvantaged** | | | | | | |
| | 0.003 | -0.013 | 0.014 | 0.002 | -0.007 | 0.021* |
| | (0.013) | (0.010) | (0.012) | (0.012) | (0.008) | (0.012) |
| Dep. var. mean | 0.411 | 0.183 | 0.325 | 0.429 | 0.169 | 0.350 |
| Observations | 53,968 | 53,968 | 53,968 | 61,222 | 61,222 | 61,222 |
| **Male** | | | | | | |
| | 0.023* | 0.012 | 0.011 | 0.005 | 0.000 | 0.005 |
| | (0.012) | (0.011) | (0.011) | (0.012) | (0.009) | (0.011) |
| Dep. var. mean | 0.454 | 0.236 | 0.325 | 0.437 | 0.193 | 0.338 |
| Observations | 60,693 | 60,693 | 60,693 | 62,571 | 62,571 | 62,571 |
| Cohort FEs | | | | Yes | | |
| Quadratic RD poly | | | | Yes | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from a regression of the outcome on $D_i$, a quadratic in the centered first-attempt score, their interactions, and cohort fixed effects, estimated within each bandwidth around the cutoff of 400. The interactions allow the quadratic trend to differ on each side of the cutoff. $D_i = 1$ if the first-attempt score is $< 400$. Columns 1–3 are Algebra I; columns 4–6 are English (Reading). Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations are from the preferred bandwidth. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth. Samples: Algebra I — TC120 (2009 Standards), cohorts 2016–2019; English — TC109, cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A13. Effect of Failing Algebra I on Mechanisms: Quadratic RD (TC120, Cohorts 2016–2019)

| | (1)<br>Att. Yr 1 | (2)<br>Att. Yr 2 | (3)<br>CTE Fam.<br>After | (4)<br>CTE Comp.<br>After | (5)<br>ACT/SAT | (6)<br>AP | (7)<br>Disab. After |
|---|---|---|---|---|---|---|---|
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | | |
| [390, 410] | 0.001 | 0.003 | 0.024** | -0.018 | 0.013 | 0.016 | -0.001 |
| | (0.002) | (0.002) | (0.010) | (0.014) | (0.016) | (0.015) | (0.004) |
| [385, 415] | -0.000 | 0.002 | 0.023*** | -0.017 | 0.011 | 0.015 | 0.004 |
| | (0.002) | (0.002) | (0.007) | (0.011) | (0.012) | (0.012) | (0.004) |
| **[380, 420]** | **-0.001** | **0.001** | **0.010*** | **-0.011** | **0.013** | **0.012** | **0.001** |
| | **(0.001)** | **(0.002)** | **(0.006)** | **(0.009)** | **(0.010)** | **(0.010)** | **(0.003)** |
| [375, 425] | -0.001 | 0.001 | 0.011** | -0.010 | 0.011 | 0.011 | -0.001 |
| | (0.001) | (0.001) | (0.006) | (0.008) | (0.010) | (0.009) | (0.003) |
| Dep. var. mean | 0.942 | 0.933 | 0.863 | 0.467 | 0.581 | 0.252 | 0.032 |
| Observations | 114,239 | 107,584 | 118,335 | 114,379 | 85,302 | 85,281 | 103,130 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | | |
| **Black & Hispanic** | | | | | | | |
| | -0.002 | -0.000 | 0.018** | -0.019 | 0.014 | -0.001 | 0.001 |
| | (0.002) | (0.003) | (0.009) | (0.014) | (0.016) | (0.012) | (0.004) |
| Dep. var. mean | 0.942 | 0.931 | 0.869 | 0.457 | 0.562 | 0.202 | 0.024 |
| Observations | 48,392 | 45,424 | 50,306 | 48,476 | 37,638 | 37,631 | 44,266 |
| **Economically Disadvantaged** | | | | | | | |
| | -0.002 | -0.001 | 0.015 | -0.008 | -0.002 | 0.011 | -0.002 |
| | (0.002) | (0.003) | (0.009) | (0.013) | (0.016) | (0.012) | (0.004) |
| Dep. var. mean | 0.930 | 0.918 | 0.865 | 0.469 | 0.482 | 0.169 | 0.024 |
| Observations | 52,268 | 48,993 | 53,968 | 52,150 | 34,708 | 34,703 | 46,773 |
| **Male** | | | | | | | |
| | 0.000 | 0.002 | 0.005 | -0.006 | 0.022 | 0.020 | -0.001 |
| | (0.002) | (0.002) | (0.008) | (0.013) | (0.015) | (0.013) | (0.005) |
| Dep. var. mean | 0.946 | 0.937 | 0.877 | 0.501 | 0.537 | 0.212 | 0.032 |
| Observations | 58,534 | 55,097 | 60,693 | 58,652 | 41,937 | 41,924 | 51,088 |
| Cohort FEs | | | | Yes | | | |
| Quadratic RD poly | | | | Yes | | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta_1 \widetilde{S}_i + \beta_2 \widetilde{S}_i^2 + \delta_1(\widetilde{S}_i \times D_i) + \delta_2(\widetilde{S}_i^2 \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $\widetilde{S}_i$ is the first-attempt score centered at 400 and the interactions allow a quadratic trend to differ on each side of the cutoff. $D_i = 1$ if first Algebra I score < 400. Cohort fixed effects (cohort 2016 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations in Panel A are from the preferred bandwidth [380, 420]. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth [380, 420]. Dep. var. mean and observations in Panel B are from the preferred bandwidth. CTE Fam. After = 1 if the student enrolled in any CTE career family in any year strictly after first Algebra I SOL. CTE Comp. After = 1 if the student attained CTE competency (`cte_competency_attainment` = Y) in any year strictly after first Algebra I SOL (0 if only N; missing = no post-SOL records, excluded). ACT/SAT = 1 if the student took the ACT or SAT in any year strictly after first Algebra I SOL; 0 if neither was taken after SOL. AP = 1 if the student took any AP exam in any year strictly after first Algebra I SOL; 0 if no AP exam was taken after SOL. Disab. After = 1 if the student (not disabled at time of test) received any new disability classification in any school year after first SOL; 0 if never classified. Sample: TC120 (2009 Standards), graduation cohorts 2016–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A14. Effect of Failing English (Reading) on Mechanisms: Quadratic RD (TC109, Cohorts 2014–2019)

| | (1) Att. Yr 1 | (2) CTE Fam. After | (3) CTE Comp. After | (4) ACT/SAT | (5) AP | (6) Disab. After |
|---|---|---|---|---|---|---|
| **Panel A: Bandwidth Sensitivity (cutoff = 400)** | | | | | | |
| [390, 410] | 0.002 | 0.012 | 0.021 | 0.011 | 0.006 | 0.004* |
| | (0.003) | (0.012) | (0.014) | (0.017) | (0.011) | (0.002) |
| [385, 415] | 0.001 | -0.001 | 0.009 | -0.005 | 0.008 | 0.004* |
| | (0.002) | (0.009) | (0.011) | (0.013) | (0.009) | (0.002) |
| **[380, 420]** | **0.001** | **0.008** | **0.015** | **-0.006** | **0.005** | **0.004**** |
| | **(0.002)** | **(0.008)** | **(0.010)** | **(0.011)** | **(0.007)** | **(0.002)** |
| [375, 425] | 0.001 | 0.010 | 0.019** | -0.003 | 0.001 | 0.004** |
| | (0.002) | (0.007) | (0.008) | (0.009) | (0.006) | (0.002) |
| Dep. var. mean | 0.918 | 0.675 | 0.521 | 0.530 | 0.129 | 0.009 |
| Observations | 107,026 | 119,545 | 105,111 | 82,747 | 82,732 | 99,110 |
| **Panel B: Subgroup Heterogeneity (BW [380, 420])** | | | | | | |
| **Black & Hispanic** | | | | | | |
| | 0.001 | 0.013 | 0.028** | -0.020 | 0.001 | 0.004* |
| | (0.003) | (0.012) | (0.013) | (0.015) | (0.010) | (0.002) |
| Dep. var. mean | 0.917 | 0.663 | 0.501 | 0.535 | 0.119 | 0.007 |
| Observations | 52,501 | 59,290 | 51,655 | 43,645 | 43,639 | 50,774 |
| **Economically Disadvantaged** | | | | | | |
| | 0.003 | -0.002 | 0.015 | -0.013 | 0.006 | 0.004* |
| | (0.003) | (0.012) | (0.014) | (0.016) | (0.010) | (0.002) |
| Dep. var. mean | 0.903 | 0.663 | 0.523 | 0.461 | 0.110 | 0.008 |
| Observations | 53,682 | 61,229 | 52,343 | 39,521 | 39,512 | 51,870 |
| **Male** | | | | | | |
| | -0.001 | 0.008 | 0.019 | 0.003 | 0.004 | 0.002 |
| | (0.003) | (0.011) | (0.012) | (0.015) | (0.008) | (0.003) |
| Dep. var. mean | 0.921 | 0.695 | 0.533 | 0.491 | 0.104 | 0.009 |
| Observations | 56,154 | 62,580 | 55,196 | 41,407 | 41,396 | 49,749 |
| Cohort FEs | | | | Yes | | |
| Quadratic RD poly | | | | Yes | | |

*Notes:* Panel A reports OLS estimates of $\tau$ from $Y_i = \alpha + \tau D_i + \beta_1 \widetilde{S}_i + \beta_2 \widetilde{S}_i^2 + \delta_1(\widetilde{S}_i \times D_i) + \delta_2(\widetilde{S}_i^2 \times D_i) + \gamma_c + \varepsilon_i$ estimated within each bandwidth around the cutoff of 400, where $\widetilde{S}_i$ is the first-attempt score centered at 400 and the interactions allow a quadratic trend to differ on each side of the cutoff. $D_i = 1$ if first English (Reading) score $< 400$. Cohort fixed effects (cohort 2014 base) included. Standard errors clustered at the school level in parentheses. Preferred bandwidth (bolded) is [380, 420]. Dep. var. mean and observations in Panel A are from the preferred bandwidth [380, 420]. Panel B reports the same specification estimated separately for each subgroup at the preferred bandwidth [380, 420]. Dep. var. mean and observations in Panel B are from the preferred bandwidth. CTE Fam. After = 1 if the student enrolled in any CTE career family in any year strictly after first English SOL. CTE Comp. After = 1 if the student attained CTE competency (`cte_competency_attainment` = Y) in any year strictly after first English SOL (0 if only N; missing = no post-SOL records, excluded). ACT/SAT = 1 if the student took the ACT or SAT in any year strictly after first English SOL; 0 if neither was taken after SOL. AP = 1 if the student took any AP exam in any year strictly after first English SOL; 0 if no AP exam was taken after SOL. Disab. After = 1 if the student (not disabled at time of test) received any new disability classification in any school year after first SOL; 0 if never classified. Sample: TC109, graduation cohorts 2014–2019. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

# B Appendix Figures

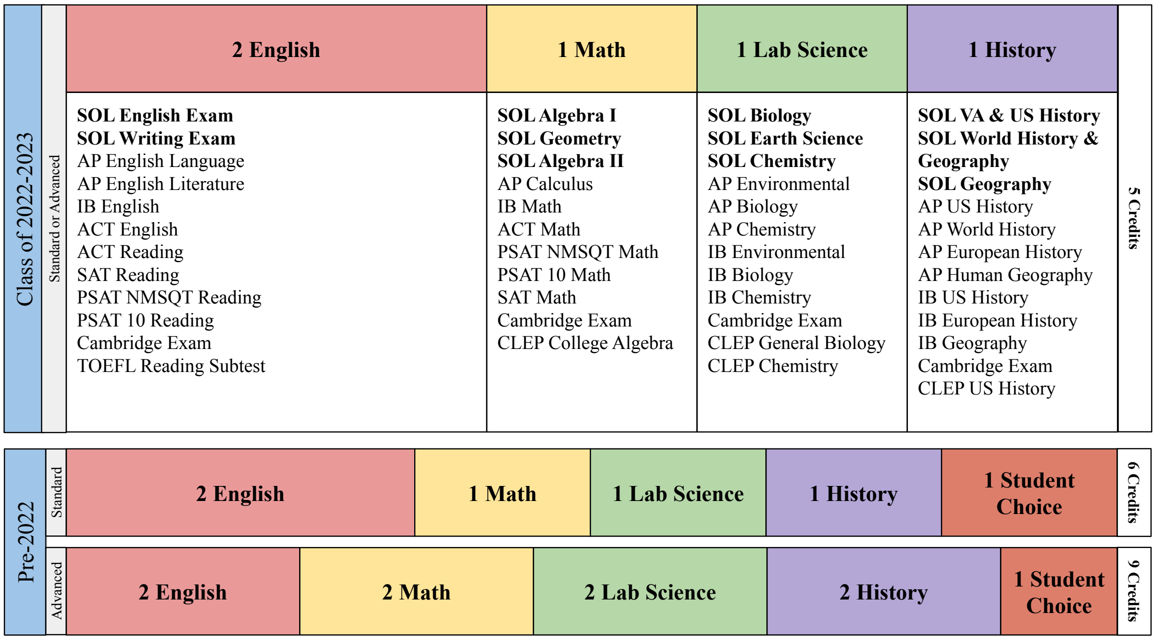


Figure A1. Verified Credits Diagram

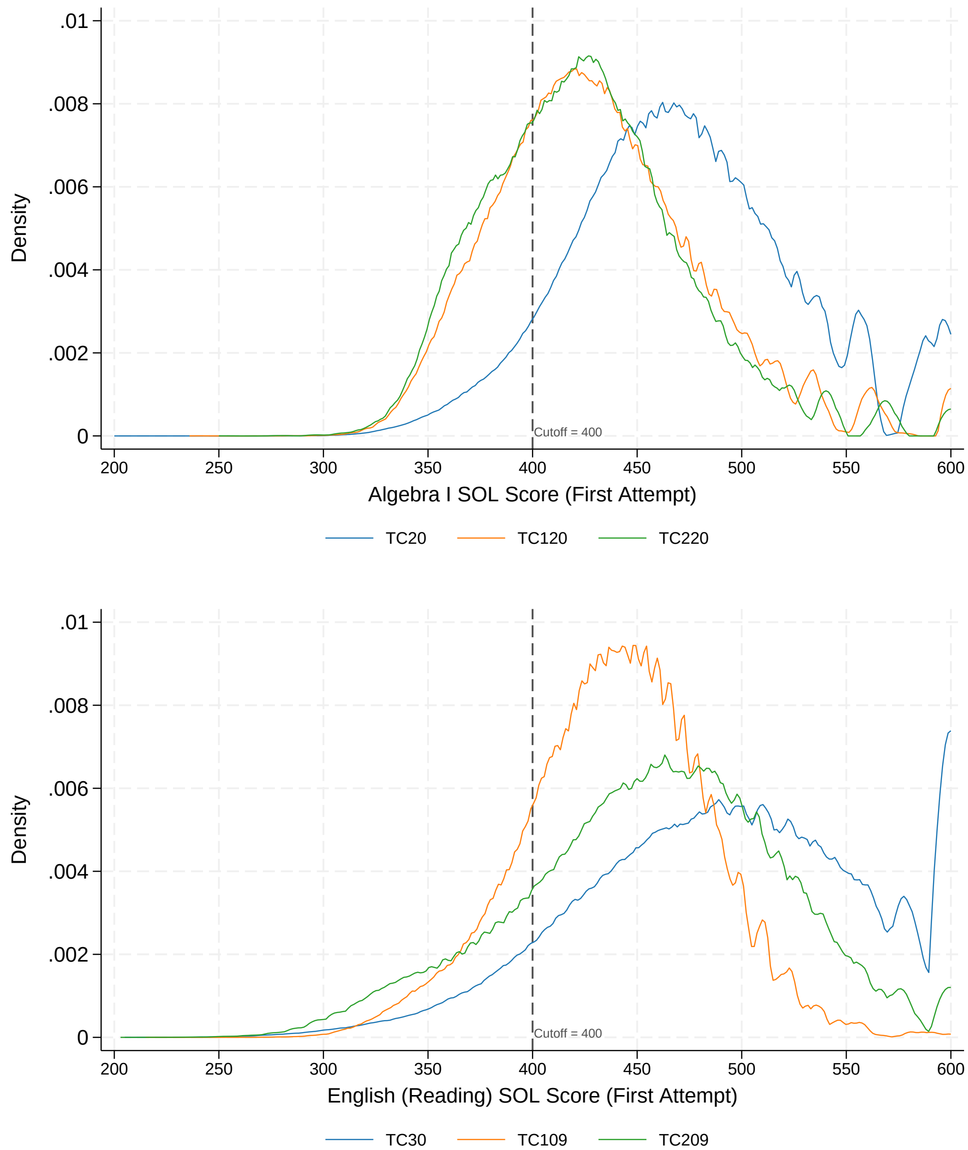


Figure A2. First-attempt score densities by test code. Top: Algebra I; bottom: English (Reading). Each curve is the kernel density of first-attempt scaled scores for one test code, with the dashed line marking the passing cutoff (400). The position of the cutoff within the score distribution differs across test codes, motivating separate analysis by test code.

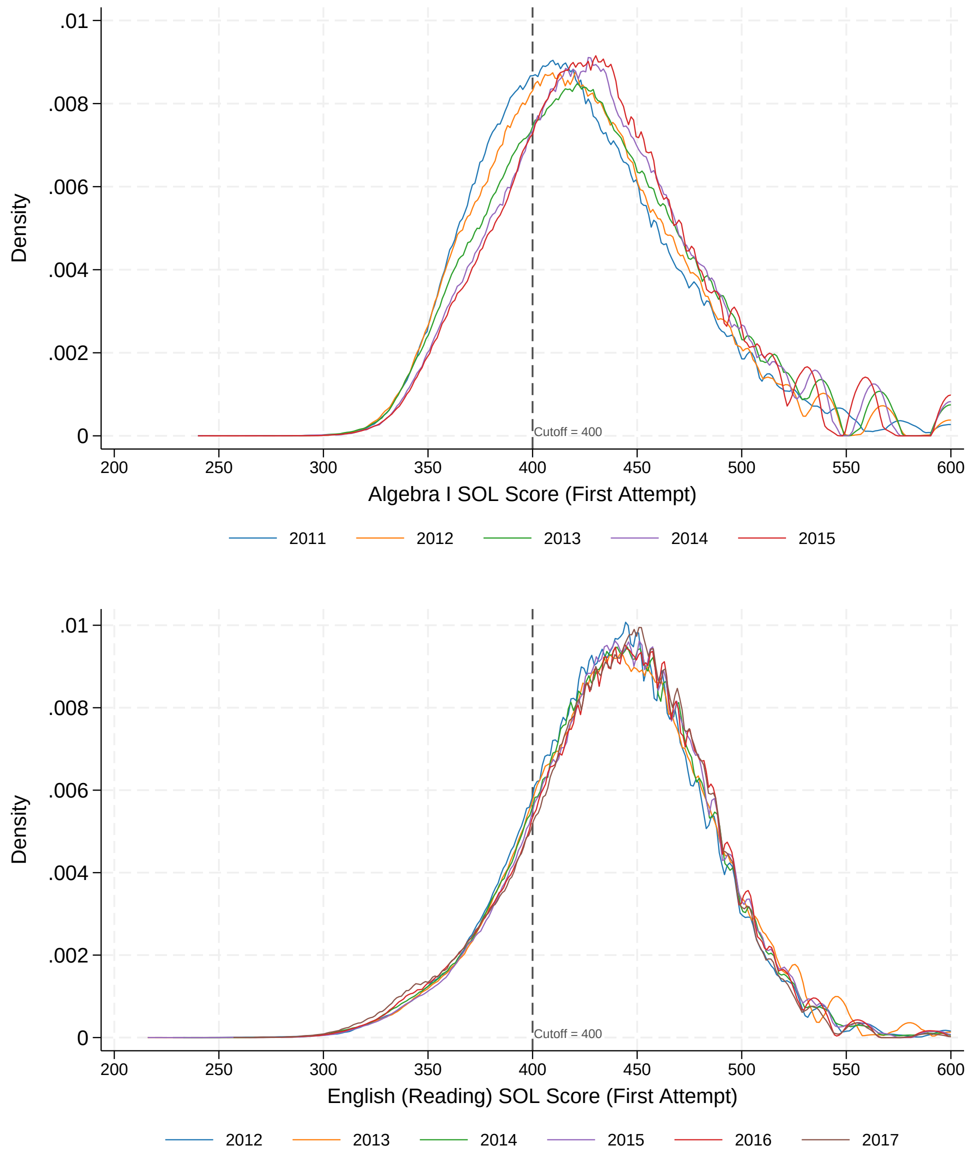


Figure A3. First-attempt score densities by test year, within test code. Top: Algebra I; bottom: English (Reading). Each curve is the kernel density of first-attempt scaled scores for one test year, with the dashed line marking the passing cutoff (400). Year-to-year distributions within a test code are similar, mitigating concern about substantial test form differences in student scores.

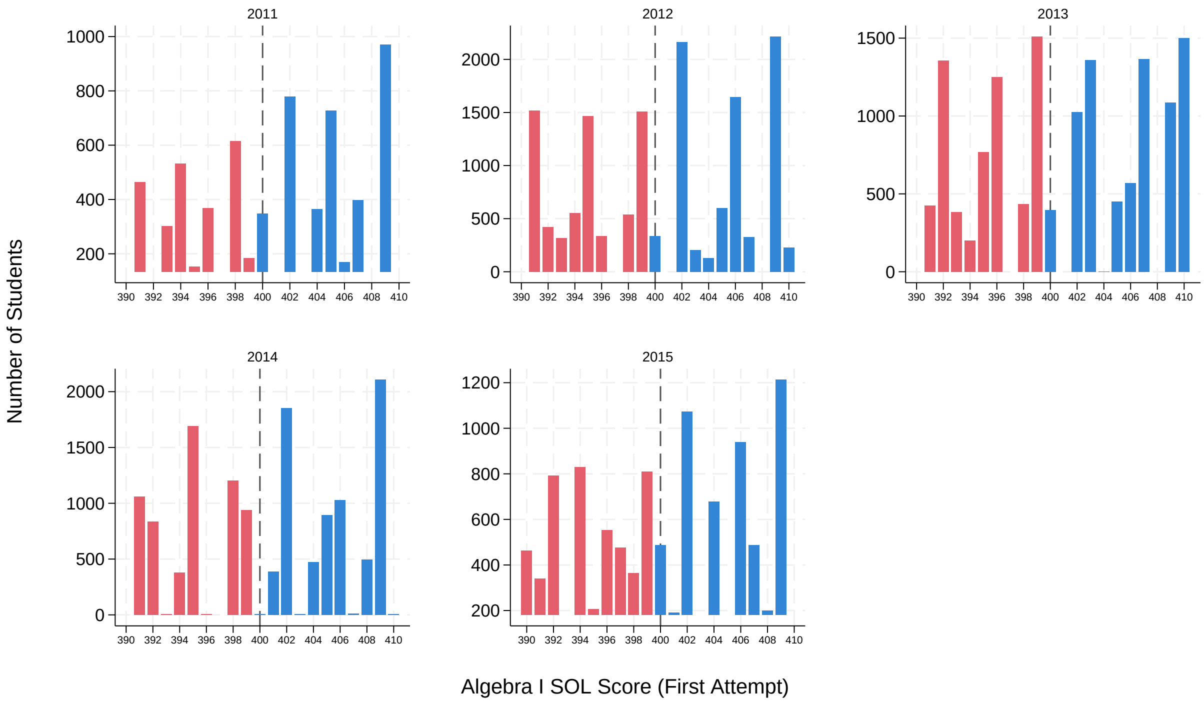


Figure A4. Score frequency in bandwidth [390, 410] by Algebra I test year. Each panel is one test year. See Figure A5 for color scheme.

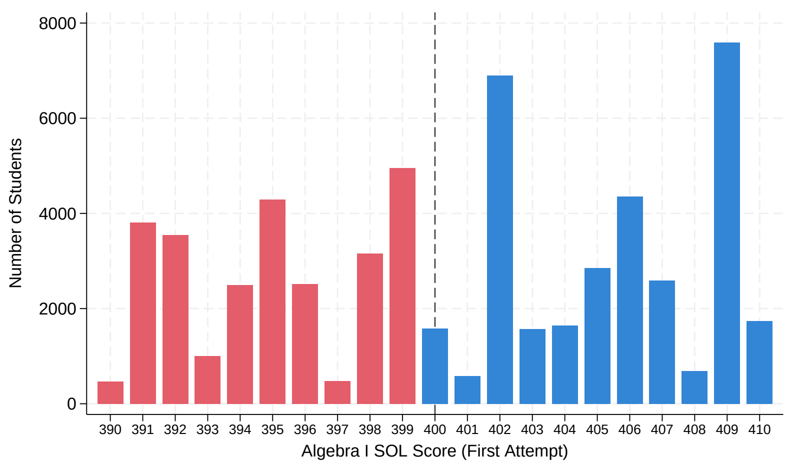


Figure A5. Score frequency in bandwidth [390, 410], pooled across test years. Sample: TC120 (2009 Standards), graduation cohorts 2016–2019. Red = below cutoff; blue = at or above cutoff. Dashed line at passing cutoff (400).

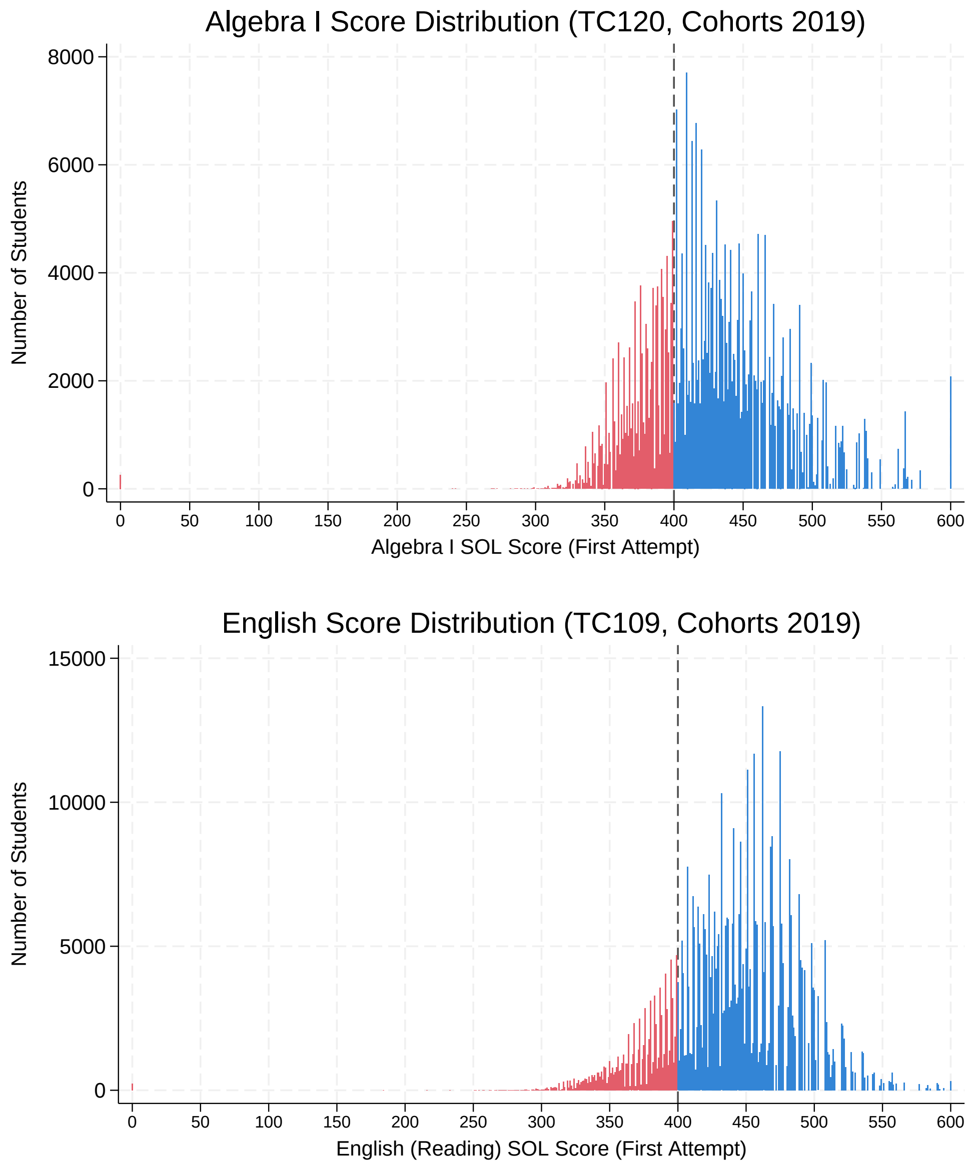


Figure A6. Full score distribution for first-attempt SOL exams. Top: Algebra I (TC120); bottom: English Reading (TC109). Each bar represents one integer score value. Red bars indicate scores below the passing cutoff (400); blue bars indicate passing scores. Sample restricted to graduation cohorts 2016–2019. The distribution shows uneven mass points across the full 0–600 scale.

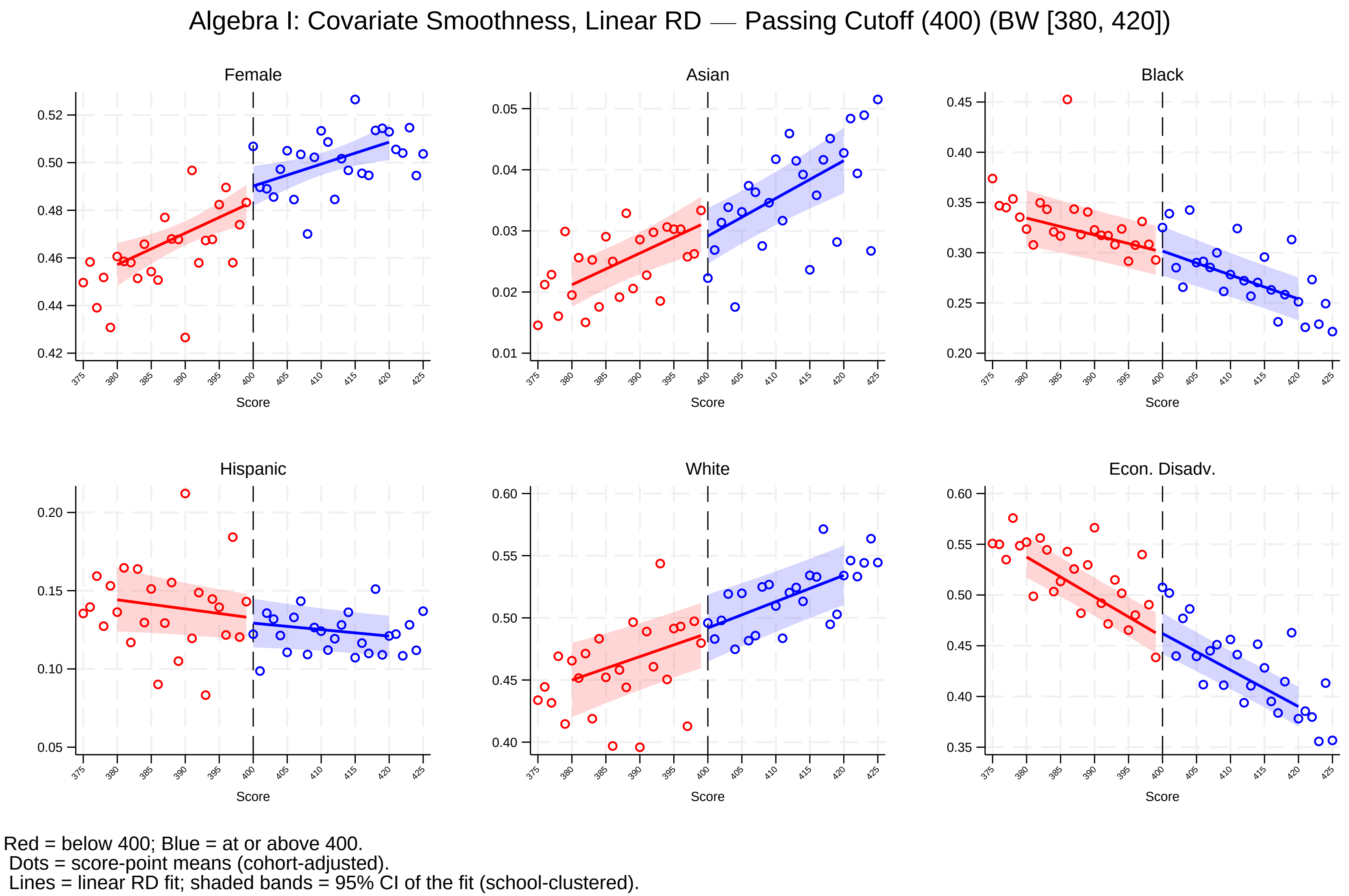


Figure A7. Algebra I: covariate balance at the passing cutoff (400). Each panel shows score-point means (cohort-adjusted) over the window [380, 420], with the linear RD fit overlaid on each side of the cutoff. Red = below 400; blue = at or above 400.

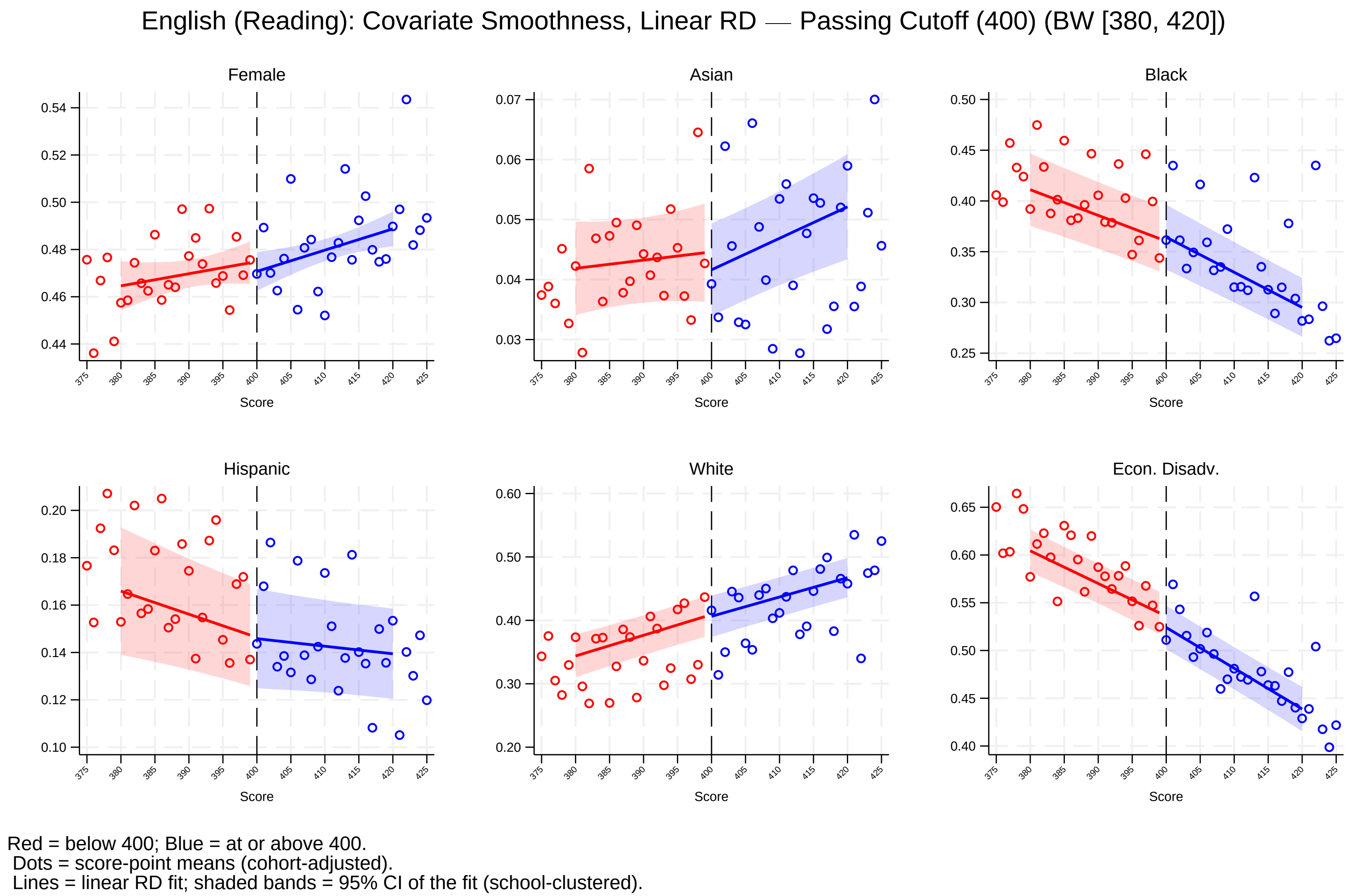


Figure A8. English (Reading): covariate balance at the passing cutoff (400). Each panel shows score-point means (cohort-adjusted) over the window [380, 420], with the linear RD fit overlaid on each side of the cutoff. Red = below 400; blue = at or above 400.

# C RD Plots

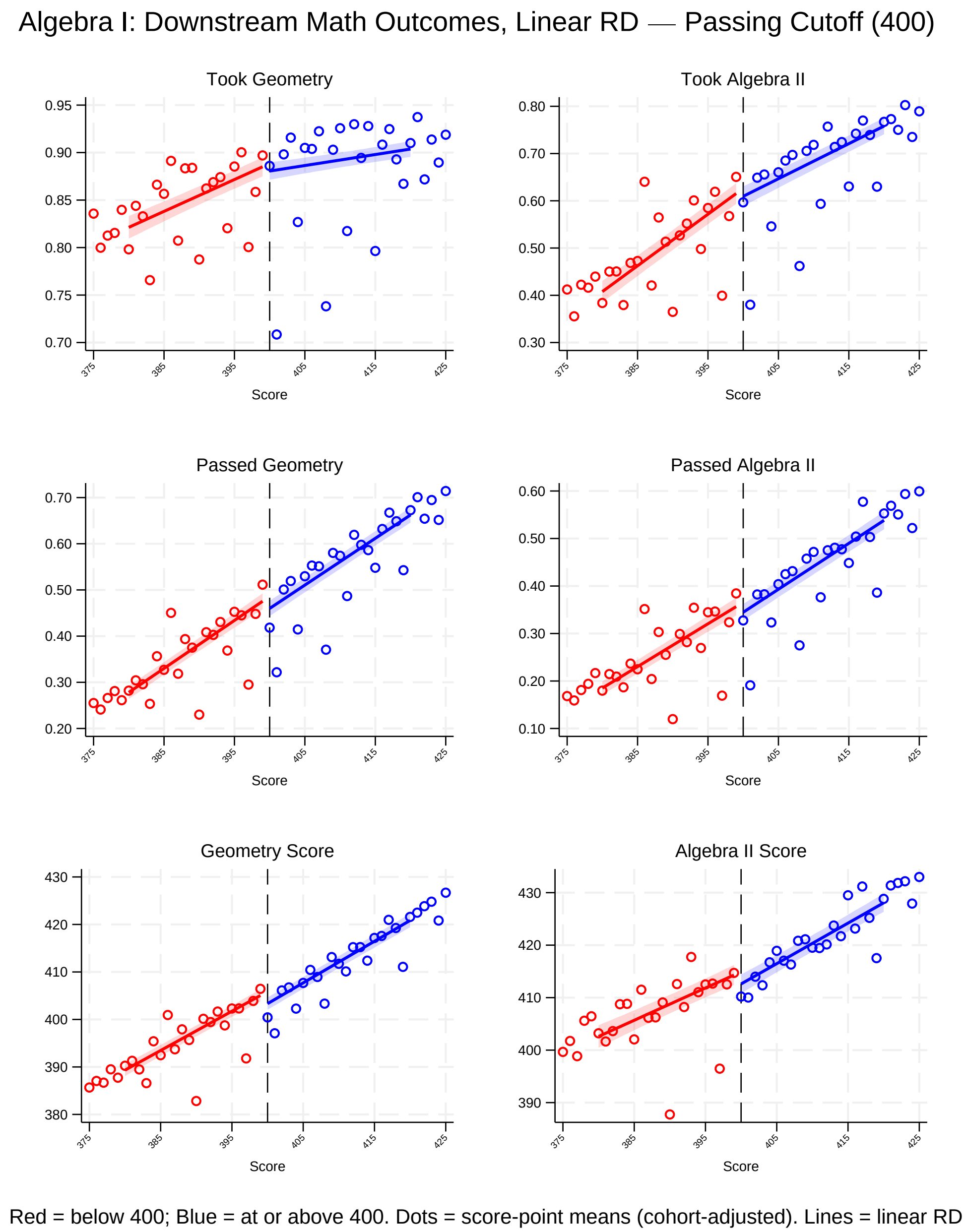


Figure A9. Algebra I: downstream math outcomes, linear RD at cutoff 400, bandwidth [380, 420]. Rows are course take-up, first-attempt passing (unconditional), and first-attempt scaled score (conditional on taking); columns are Geometry and Algebra II, mirroring Table 2. Score panels are on the 0–600 scaled-score axis. Dots show score-point means (cohort-adjusted) over the window [375, 425]. Lines show the linear RD fit with separate slopes on each side of the cutoff. Shaded bands are 95% CIs of the fit from school-clustered regression. Red = below 400; blue = at or above 400.

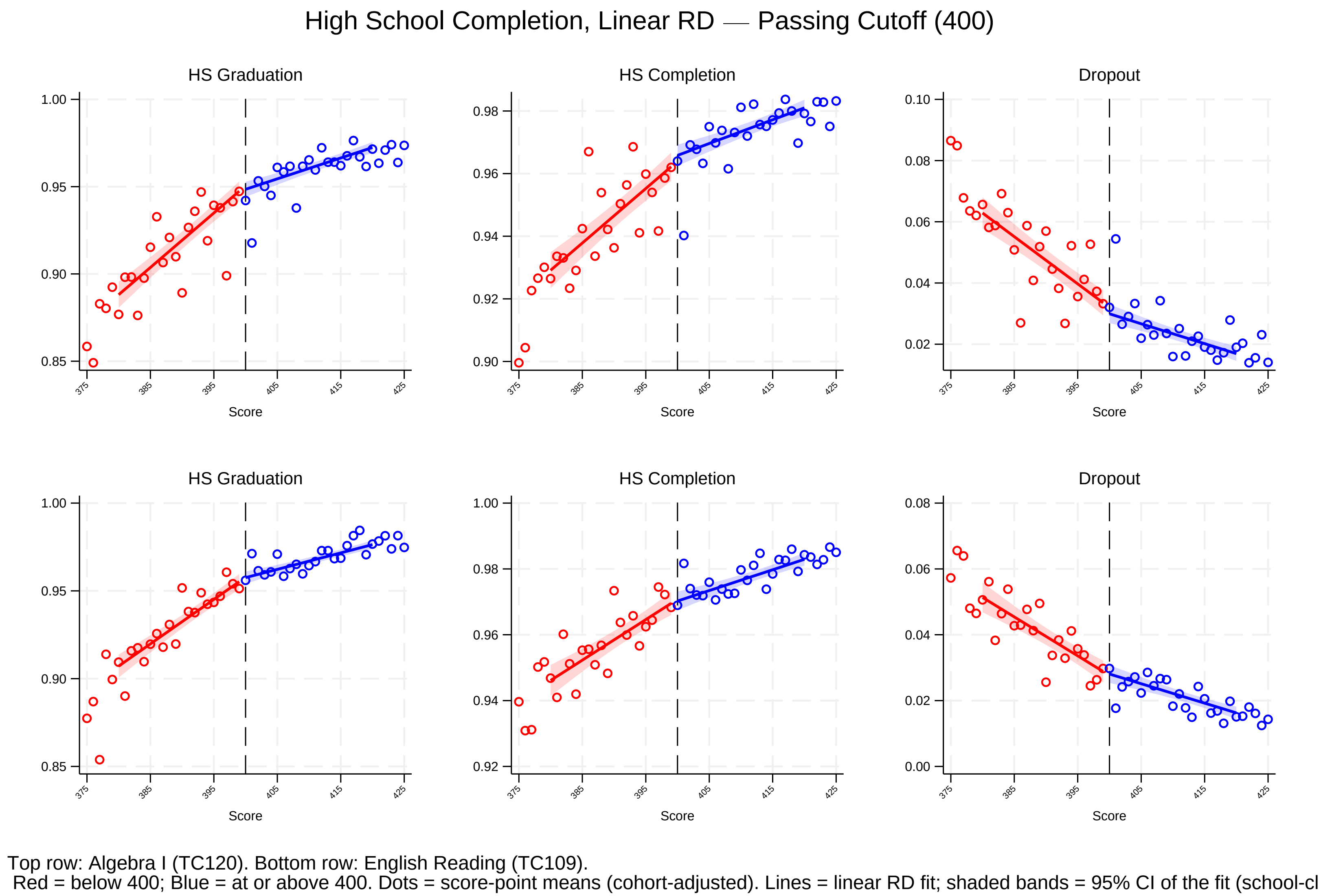


Figure A10. High school completion, linear RD at cutoff 400, bandwidth [380, 420]. Top row: Algebra I (TC120); bottom row: English Reading (TC109). Columns are high school graduation, high school completion, and dropout, mirroring the High School Completion columns of Table 3. Dots show score-point means (cohort-adjusted) over the window [375, 425]. Lines show the linear RD fit with separate slopes on each side of the cutoff. Shaded bands are 95% CIs of the fit from school-clustered regression. Red = below 400; blue = at or above 400.

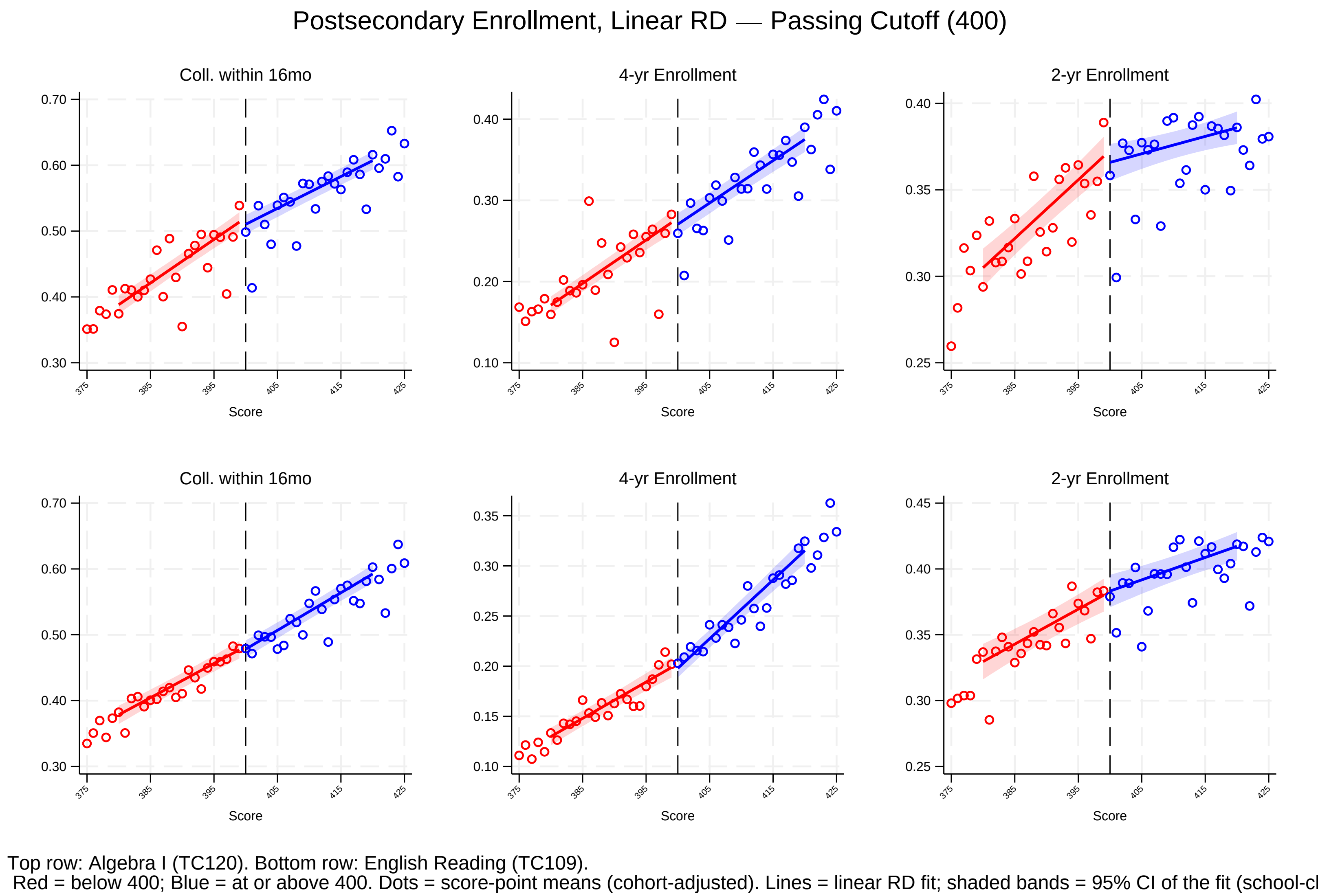


Figure A11. Postsecondary enrollment, linear RD at cutoff 400, bandwidth [380, 420]. Top row: Algebra I (TC120); bottom row: English Reading (TC109). Columns are college enrollment within 16 months, 4-year enrollment, and 2-year enrollment, mirroring the Postsecondary Enrollment columns of Table 3. Dots show score-point means (cohort-adjusted) over the window [375, 425]. Lines show the linear RD fit with separate slopes on each side of the cutoff. Shaded bands are 95% CIs of the fit from school-clustered regression. Red = below 400; blue = at or above 400.

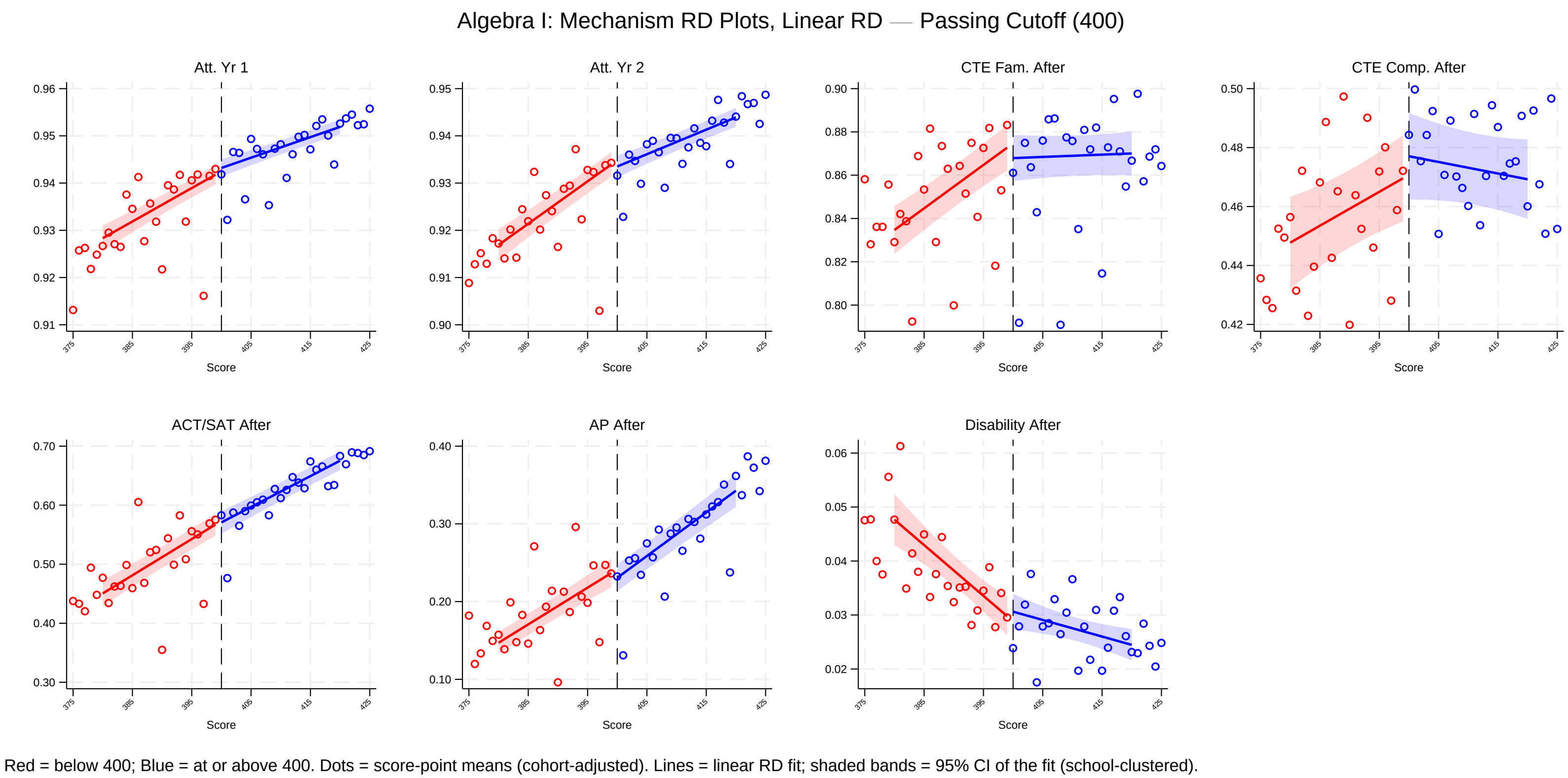


Figure A12. Algebra I: mechanism RD plots, linear RD at cutoff 400, bandwidth [380, 420]. Panels mirror the Algebra I columns of Table 4: attendance in the year of the exam and the year after, CTE course-family and competency enrollment after failing, ACT/SAT and AP participation after failing, and new disability classification. Dots show score-point means (cohort-adjusted); lines show the linear RD fit with separate slopes on each side of the cutoff; shaded bands are 95% CIs from school-clustered regression. Red = below 400; blue = at or above 400.

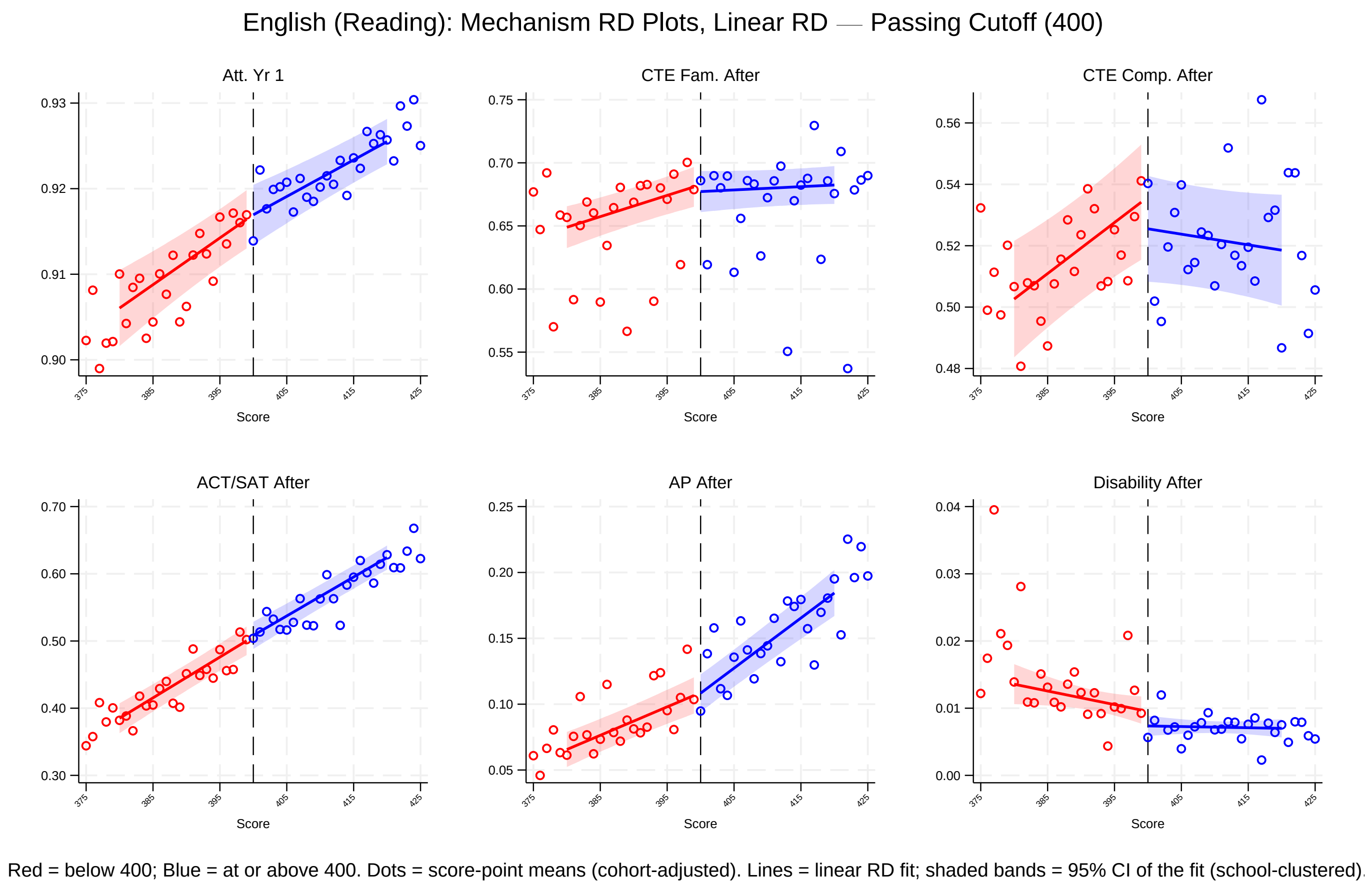


Figure A13. English (Reading): mechanism RD plots, linear RD at cutoff 400, bandwidth [380, 420]. Panels mirror the English columns of Table 4: attendance in the year of the exam, CTE course-family and competency enrollment after failing, ACT/SAT and AP participation after failing, and new disability classification. Dots show score-point means (cohort-adjusted); lines show the linear RD fit with separate slopes on each side of the cutoff; shaded bands are 95% CIs from school-clustered regression. Red = below 400; blue = at or above 400.