Rules versus Discretion in the Use of Educational Testing

David Figlio, University of Rochester and NBER[†]

Umut Özek, RAND Corporation

This version: December 8, 2025

PRELIMINARY AND INCOMPLETE: PLEASE DO NOT CITE OR DISTRIBUTE WITHOUT PERMISSION

Declarations of interest: None

[†] The views, opinions, findings, conclusions, and recommendations contained herein are the authors' alone and not those of their employers or their research sponsors, clients, or grantors. The authors thank the anonymous school district for providing the data used in the analysis, and for providing useful feedback. Note that the views expressed are those of the authors and do not necessarily reflect those of the anonymous district or the institution to which the authors are affiliated.

Rules versus Discretion in the Use of Educational Testing

Abstract: In 2015, Florida changed its rules regarding student remediation in middle/high school from a system in which remediation was determined relatively strictly based on test score proficiency cutoffs to one with considerable educator discretion. We make use of difference-in-difference-in-difference and difference-in-difference-in-discontinuity designs and across students with dramatically different predicted treatment intensities across regimes, and find that students tend to have better long-run college outcomes (attendance, persistence, and success) when educators have discretion versus when they must follow proficiency-based rules. We also exploit a year in which, due to a glitch, no new test score information was available to provide evidence that our results are due to relative benefits of discretion versus rules, as opposed to "less of a bad treatment", as we find that remediation is apparently beneficial for many students whom educators are most likely to remediate based on non-proficiency-related factors.

Keywords: rules versus discretion; educational interventions; eligibility cutoffs; human capital

Rules versus Discretion in the Use of Educational Testing

1. Introduction

Examples abound regarding the use of discrete classifications of continuous variables to make decisions about resource allocation, service eligibility, or treatment assignment.[2] But there is tension between reliance upon rules versus allowing for professional discretion. There are many arguments for the use of strict rules for resource allocation and treatment assignment: The use of rules generates more transparent, predictable decision-making, establishes clear criteria that resources are allocated as intended, and reduces the likelihood that bias, discrimination, favoritism, or corruption will influence outcomes.[3] On the other hand, discretion allows professionals with deep, multifaced understanding about a specific case to holistically determine the best treatment for the case at hand,[4] and reliance on arbitrary thresholds can lead to inefficient outcomes.[5]

It's not obvious that human judgment is better than algorithms or proficiency tests for many decisions: There is evidence that managers who hire low-skilled workers against testing

[2] A handful of examples of this phenomenon include: Birth weight thresholds, such as very low birth weight (<1500 grams) or extremely low birth weight (<1000 grams) often automatically trigger different levels of neonatal intensive care (as well as different Medicare Severity Diagnosis-Related Groups and consequent levels of hospital reimbursement), the use of mechanical ventilation and surfactant administration, and eligibility for early intervention services under Part C of the Individuals with Disabilities Education Act. Estimated Glomerular Filtration Rate thresholds determine priority on waitlists for the allocation of kidney transplants. Some states and counties use COMPAS risk scores to determine conditions for pretrial release, probation intensity, or sentencing. Section 8 Housing Choice Vouchers rely on income bands based on percentage of area median income to determine families' access to and level of rental assistance. Air quality public health alerts are triggered by crossing discrete thresholds of measured pollutant concentration. Many jurisdictions allocate social service funds based on current or past income levels. Some higher education institutions and systems automatically place students into remedial courses if they are below a standardized test score threshold, and some states and localities mandate K-12 grade repetition or student remediation if students score below a certain level on standardized test scores.

[3] Sometimes rules are developed explicitly for equity considerations, such as those that limit parental decision-making. See, e.g., Ricks, 2024, for a recent example regarding school starting age restrictions.

[4] This potential tension is apparent in cases such as teacher-graded high-stakes exams, where there are numerous examples of educators manipulating high-stakes exam scores to move some students into preferred treatments (Borcan, Lindahl, and Mitrut, 2017; Dee et al., 2019; Diamond and Persson, 2016).

[5] Carril (2022), for instance, shows that public procurement rules with arbitrary thresholds introduce compliance costs that outweigh the benefits from waste reduction in the procurement setting. Busso et al. (2024) find that requiring schools to hire teachers with a certain test score reduces student outcomes.

recommendations tend to make worse hires on average (Hoffman, Kahn, and Li, 2018), universal screening for gifted education in elementary school leads to more low-income and minority students being identified as gifted (Card and Giuliano, 2016), and "red flag" drug prescriptions against treatment guidelines tend to lead to worse youth mental health outcomes (Cuddy and Currie, forthcoming). Meanwhile, objective test-based information can help managers update their priors about employees (Rockoff et al., 2012) and machine learning can enhance professional decision-making (Kleinberg et al., 2018). All told, there is ample evidence to suggest both that professional judgment is imperfect and yet that reliance on "objective" information alone is also imperfect.

While there are several excellent recent studies comparing downstream outcomes when professionals exercise discretion versus following guidelines (e.g., Cuddy and Currie, forthcoming; Hoffman, Kahn, and Li, 2018), we know of no studies that compare outcomes across *policy regimes* where discretion is permitted versus where roles are (generally) strictly enforced.[6] To fill this gap in the literature, we exploit a natural experiment to investigate the relative benefit of adhering to strict rules versus allowing educator discretion in the use of standardized tests. Florida's test-based middle/high school student remediation policy shifted in 2015 from a strict rule-based policy – in which students scoring below proficient on their English Language Arts (ELA) exam were required to take a "double dose" of ELA courses, both a standard course and a remedial course, the next year – to one that provided the same proficiency guideposts for educators but that also permitted considerable educator discretion.[7] Remediation

[6] The closest study we have found to this question is Oh (2025), who studies the degree to which judge discretion in the assignment of cases to defense attorneys promotes corruption. She finds that when judges can no longer assign cases to defense attorneys, attorneys' contributions to judicial election campaigns fall and defendants are less likely to be convicted.

[7] Importantly, for the grades that we are considering – eighth and ninth grade exams – there were no other major consequences for students (e.g., grade retention) of scoring on one side of this proficiency threshold or the other, either before or after the policy change. There were some potential consequences for schools, as Florida's school

probabilities are strongly affected by what side of the proficiency threshold a student is on, regardless of policy regime, but treatment intensity – the likelihood that a student is remediated conditional on being just below the proficiency line – is dramatically lower in the discretion regime than in the rule regime.

We make use of detailed student-level information from an anonymous Florida school district, one of the ten largest in the United States, to predict which students would and would not be remediated in the discretion regime, and then compare the postsecondary outcomes of extremely similar students, grouped by their remediation probabilities, in the rule regime versus the discretion regime. We investigate whether students extremely close to the ELA proficiency margin – and who scored below the bright-line proficiency threshold for the first time in either of these regimes – have better postsecondary outcomes (attendance, persistence, and graduation) when schools are directed to follow a strict rule versus when educators have discretion regarding whether to remediate students. As observationally identical students (measured along a range of dimensions, such as past grades, attendance, behavior, and test scores) who score just above the bright-line proficiency threshold are almost never remediated – even among those predicted to have a high probability of remediation in the discretion regime if they scored a couple of points lower on the standardized test – this lends itself to a triple-difference or difference-in-difference-in-regression discontinuity design. We estimate both types of models, comparing the difference in the discretion versus rule regimes in the difference between students predicted to be least likely to be remediated in the discretion regime to those most likely to be remediated in the difference in the difference between students just below the threshold for the first time (and

accountability system at the time assigned letter grades to schools based in part on the fraction of students who scored above the proficiency level. This remained unchanged across the two policy regimes.

therefore sometimes/generally remediated, depending upon regime) and those just above the threshold (and therefore almost never remediated).

We find evidence that students were being over-remediated in the strict rule regime versus the educator discretion regime. We estimate that students with low probabilities of remediation in the discretion regime have better college outcomes – higher probabilities of enrollment, persistence, and two-year degree graduation rates – than they would have experienced in the rule regime. The college graduation results are operating through increased two-year graduation rates rather than four-year graduation rates, but we find no evidence of crowd out.

We then seek to understand the intermediate outcomes along the way that might help to shape these findings. This could come about through a combination of educator, student, and parental behaviors: Assignment to remediation could lead to tracking in other subjects (see, e.g., Figlio and Özek, 2024) and scoring below a salient threshold could influence teenagers' subsequent decision-making (see, e.g., Papay, Murnane, and Willett, 2014). We therefore study how scoring just below the ELA proficiency threshold in the rule versus discretion regimes affects different groups of marginally-performing students in terms of their subsequent course-taking (including their likelihood of taking college-preparatory or advanced-placement classes), their attainment of a set of classes that would better prepare them for college, the prior performance levels of students' classroom peers, and (when possible) the quality of teachers, as measured by value-added, that students experience. We further investigate whether contextual factors of the school – the socio-economic status and racial composition of the student body, as well as the dispersion in prior performance of the student body – affect these relationships. We provide evidence that the beneficial consequences of educator discretion for marginal students

may be due in part to higher-scoring peers in both ELA and other core courses, as well as more advanced college-preparatory course-taking.

We also find evidence of heterogeneous effects of discretion versus rules by race, socio-economic status, gender, home language, and school attributes. We show that the estimated effects of discretion are positive across all subgroups, but are not uniform across groups or always statistically distinct from zero. We find that discretion apparently improves two-year degree attainment the most for Hispanic students, non-native English speakers, and male students, while it apparently improves four-year degree attainment the most for White students, native English speakers, students from relatively affluent families, and male students. Two-year degree attainment results are strongest in majority nonwhite schools, relatively low-socio-economic status schools, and schools where students tend to have higher test scores. Meanwhile, we find relatively weak evidence of differential treatment of students under the discretion regime.

Given that we find evidence that allowing for educator discretion improves longer-run outcomes for students whom educators were relatively less likely to remediate, one might reasonably wonder if *no remediation* is better than educator discretion – that is, are our findings less about the relative value of "discretion versus rules" than about "less of a bad treatment"? An unusual circumstance in this case allows us to also derive some suggestive evidence regarding what educators would do in the absence of the bright-line proficiency threshold. Due to a significant fluke in the implementation of Florida's standardized test in 2015, the tests were administered later than usual and scores were not reported to educators until mid-fall of the next academic year. Therefore, educators had to make course placement decisions for the 2015-16 academic year with all the data they typically had at their disposal (including the myriad factors

we use to estimate discretion-regime remediation probabilities) but not the 2015 test scores – or where students scored relative to the proficiency threshold.

Two features characterize educator remediation placement decisions in 2015, the year when they were forced to make placement decisions without new test data. First, educators assigned almost no new students to remedial coursework, regardless of which side of the (unobserved) line students scored. Second, while in both the rule regime and the discretion regime educators almost always immediately ceased remediation for previously-remediated students who scored just above the bright-line proficiency threshold, in the 2015-16 academic year three-quarters of previously-remediated students *continued to be remediated*. Therefore, while it is impossible to study what happens to students who are newly-remediated in environments with and without observed test scores, we can study what happens to students whom educators elected to continue to remediate in the no-test environment who would almost certainly not have been remediated in either *testing-visible* regime.

We therefore carry out two sets of suggestive analyses in an attempt to provide some answers to the question of whether our results better reflect the benefits of discretion versus rules or the "less of a bad treatment" story: We first compare students in the discretion regime, with varying degrees of likelihood of being remediated under educator discretion, to observationally identical sets of students in the year where there was no new testing information. Here, we find that similar students don't fare appreciably better or worse under discretion than the no new test information condition. We then, however, investigate whether students already flagged for remediation but newly deemed proficient – but whose proficiency status was unknown, at the time of course placement, to students, families, and educators – fared better or worse than they would have in the rules regime if their new measured proficiency status were known. We

compare students whom educators would have been relatively likely or relatively unlikely to remediate under the discretion regime based upon non-test factors – the same predictions we used in our main analyses – and find that the students whom educators would have been most likely to remediate under discretion (but for the test data and proficiency measure) benefited the most in the year in which *remediation continuation decisions* were based on educator discretion without the benefit of test score or proficiency data. We view this as evidence that the correct level of remediation is not zero, and that the strong use of proficiency thresholds as a remediation-cessation guidepost might have ended remediation prematurely for some students who would have benefited from continued remediation. This also speaks to the potential benefit for the use of educator discretion in a longer-run remediation system in which proficiency guideposts are present but less salient.

This paper, therefore, demonstrates an important role for professional discretion in an environment in which professionals have a lot of private information about the individuals about whom they are making decisions. The presence of private information could be one explanation for the differences between our results and the findings of Hoffman, Kahn, and Li (2018), where hiring managers likely do not have high-quality private information about job candidates to contravene the recommendations of job testing.

2. Policy Background

Florida's middle and high school remedial course policy was first enacted in 2004 and later expanded in 2006 (s.1003.415 in 2004 and s.1003.4156 in 2006 for middle school; s.1004.428 in 2006 for high school). The policy required students who scored in the lowest two achievement levels (out of five) on the standardized, statewide reading and mathematics

assessments to take a remedial course in that subject along with the regular course. There are exemptions to this requirement: For example, students may be exempt from taking an additional remedial course in ELA if they have higher remediation needs in math. Similarly, low-performing students who do not have intervention needs in the areas of foundational reading skills (e.g., decoding, fluency) or students who have consistently scored above the proficient level in the past may be exempt from ELA remediation. That said, the policy significantly increased remedial course-taking in middle and high school in the state (e.g., Figlio and Ozek, 2024).

The primary objective of the policy was to improve student achievement in the remediation subject by providing additional instructional resources to students. In particular, the policy adopted a multi-pronged approached: (1) it required districts to provide additional instruction time to low-performing students in the remediation subject; (2) it required remedial courses to be staffed by teachers with evidence of success (as determined by the district); (3) required districts to provide differentiated support based on the academic needs of students in the remedial schedule; and (4) provide remedial courses in classrooms with adequate infrastructure (class size, materials). Recent evidence on this policy present evidence supporting the fidelity of implementation in middle school and high school, showing that students in the remedial schedule were indeed assigned to higher value-added teachers and spent more time in the classrooms of remediation subject (Figlio and Özek, 2024; Özek, 2021, Özek, 2025). There is also evidence suggesting positive long-term effects of middle school remedial courses in ELA on postsecondary outcomes (Özek, 2021) while negative effects on postsecondary degree attainment of high school remedial courses (Özek, 2025).[8]

[8] There are several possibilities that might explain the discrepancy between the middle school and high school effects. For example, it is possible that being held back (in terms of course content) has a more adverse effect in later

This requirement was eliminated in 2015 (HB 7069) and the state left the remediation decision to the individual school districts.[9] We call the period up to 2015 the *rule regime*. Aside from this change in state policy, there were issues with the implementation of the new state test (named Florida Standards Assessment, or FSA) in 2015 and test scores were not available to educators until they were independently verified in Fall 2015. As such, in the summer before the 2015-16 school years, educators had to make remediation decisions for middle and high school students without test scores and in the absence of a state mandate. We call the 2015-16 school year the *no test regime*. After 2015-16, educators made the remediation decisions with test scores, but without the state mandate. We call this period the *discretion regime*. In our analysis, we focus on ELA remediation because remedial course-taking in math was uncommon in the LUSD during this time frame: Only 5 percent of all high school students took a remedial math course during the time frame that we examine (compared to roughly 25 percent for ELA).

3. Data

In this study, we make use of detailed longitudinal, student-level school records from a LUSD in Florida that cover all students enrolled in grades 6–12 between 2011–12 and 2022–23 school years. LUSD is one of the largest school districts in the United States, serving more than 200,000 students in grades K-12 from diverse backgrounds. These data include information on ELA and mathematics scores of all tested students; demographic information on students, such as race and gender; free-or-reduced price lunch (FRPL) eligibility, English learner status,

grades as students have less time to catch up, similar to grade retention policies (Jacob and Lefgren, 2004; Greene and Winters, 2007; Greene and Winters, 2012; Mariano and Martorell, 2013; Schwerdt et al., 2017; Figlio and Ozek, 2020; Hwang and Koedel, 2023; Mumma and Winters, 2023; Jacob and Lefgren, 2009; Eren et al. 2017; Mariano et al, 2018; Larsen and Valant 2018; Eren et al. 2022). It is also possible that being labeled as low performing has a more detrimental stigma effect on older students, leading to student disengagement from schooling.

[9] https://www.fldoe.org/core/fileparse.php/7506/urlt/Information-on-House-Bill-7069.pdf, accessed on 9/1/2025.

exceptional/special education status, country of birth, language spoken at home, student age, schools attended, student disciplinary problems, and attendance. More importantly for the purposes of this study, we also observe detailed course enrollment records (including unique course identifiers and name, which we use to identify remedial and advanced courses, and teacher identifiers). [10] These K–12 records are linked to NSC records for students who graduated from a high school in the district between 2013 and 2022 (and follow them until June 2024) which allow us to observe postsecondary enrollment, college selectivity and completion for students.

In our analysis, we focus on 9th and 10th graders between 2012-13 and 2017-18 school years, which gives us 3 years of rules regime and 2 years of discretion regime. We chose this timeframe in order to observe postsecondary completion for all students in the analytic sample. We exclude 11th graders from the analysis since the remediation cutoff on the 10th grade ELA test (i.e., achievement level 2/3 threshold) was also used as a criterion for high school graduation in Florida; and 12th graders because remedial courses were not as common for high school seniors.

Figure 1 and Figure 2 present the likelihood of taking a remedial ELA course around the remediation cutoff (proficiency cutoff based on prior year's ELA test score) for students who had never been flagged for remediation in the past two years (i.e., never-flagged students) and for students who were taking a remedial ELA course in the previous school year (i.e., previously-remediated students) under the three regimes (Online Appendix Figure 1 and Online Appendix Figure 2 present the results from the same analysis for each school year separately while Online

[10] Test scores include Florida Comprehensive Assessment Test (FCAT) scores in reading and mathematics for all students in Grades 3–10 until the 2011–12 school year; FCAT 2.0 scores in reading for Grades 3–10 and in mathematics for Grades 3–8 between 2012–13 and 2014–15; and Florida Standards Assessment (FSA) scores in reading for Grades 3–10 and in mathematics for Grades 3–8 since 2014–15.

Appendix Figure 3 presents the results from year-by-year analysis for all students). Online Appendix Figure 4 demonstrates that there is no apparent manipulation around the cutoff, a fact also shown by Özek (2025).

There are several takeaways. First, the findings presented in Panel (A) of Figure 1 and Figure 2 suggest that the rule was strongly enforced under the rule regime. For example, nearly 75 percent of all never-flagged students and almost 90 percent of previously remediated students within 10 points below the cutoff were placed in the remedial schedule under the rule regime. In contrast, in Panel (C) of each graph, we find considerable discretion under the discretion regime, especially for the never-flagged students: only a third of these students within 10 points below the cutoff were placed in remediation under the discretion regime (about 60 percent of previously remediated students right below the cutoff were placed in remediation under the discretion regime). Interestingly, when educators did not have access to test scores, as shown in Panel (B) of Figure 1 and Figure 2, they chose to retain the status quo: They mostly did not place students in remediation for the first time and generally placed previously remediated students in the remedial ELA schedule again.

4. Empirical Strategy

Our primary research question examines whether students right below the ELA remediation cutoff (within 10 points below the cutoff) have better outcomes under the rule versus discretion regime. Table 1 presents the characteristics of students within 10 points below and 10 points above the remediation cutoff overall (first two columns), under the rule regime (columns three and four), under the no test regime (columns five and six), and under the discretion regime (last two columns). Our analytic sample is highly diverse with White, Hispanic, and Black

students nearly equally represented; 65 percent of the students eligible for subsidized meals, and 31 percent speaking a language other than English at home. As expected, students below the cutoff have worse academic outcomes in prior year with higher rates of absences, disciplinary incidents, worse performance on math tests, and more likelihood of having taken a remedial course in the previous year compared to students above the cutoff. Yet, importantly for the purposes of our study, this difference is similar under the three regimes.

In our main analysis, we focus on never-flagged students for whom we observe a more substantial shift in remediation placement between rule and discretion regimes if they fall below the remediation cutoff (as shown in Figure 1). Using these students (never-flagged students within 10 points below the remediation cutoff) in the discretion regime, we first regress the remediation indicator on student baseline attributes such as course-taking history (indicators for course identifiers in the past two years), past attendance, behavior, and test score history. Using the estimated coefficients on these predictors, we then obtain the predicted likelihood of remediation for never-flagged students above the cutoff and under the rule and no test regimes, and identify low-likelihood students (predicted probability lower than 0.2), medium likelihood students (predicted probability between 0.2 and 0.4), and high likelihood students (predicted probability higher than 0.4) based on the distribution of predicted probabilities (these cut-points roughly correspond to the terciles of the distribution).

Table 2 presents the realized remediation probabilities for low-, medium, and high-predicted probability students under the three regimes, revealing significant discrepancies for never-flagged students under the rules versus discretion regimes based on whether they fall above or below the cutoff. For example, under the rule regime, never-flagged students face similar remediation rates regardless of their predicted remediation likelihood: If they score below

the cutoff, 67 percent of never-flagged, low-likelihood students are placed in remediation under the rule regime while the same number is 73 percent for high-likelihood students; If they score above the cutoff, only 0.6 percent (1.5 percent) of never-flagged low-likelihood (high-likelihood) students are remediated.

In contrast, never-flagged students face very different consequences under the discretion regime if they score below the cutoff: Only 6 percent of never-flagged, low-likelihood students are placed in remediation under the discretion regime whereas the same rate is 61 percent for high-likelihood students if they score below the cutoff. If they score above the cutoff under the discretion regime, they are rarely remediated regardless of their predicted remediation likelihood (0.8 percent for low-likelihood students versus 2.8 percent for high-likelihood).

In our identification strategy, we leverage this variation in the "remediation shock" in treatment intensity (i.e., the difference between realized versus predicted remediation likelihood) among never-flagged students, comparing postsecondary outcomes for low-predicted likelihood students versus those with high-predicted likelihood, below the cutoff versus above the cutoff, and under the discretion regime versus the rule regime in a simple difference-in-difference-in-differences (DDD) design and a difference-in-difference in regression discontinuity (DiD in RD) design that we estimate using a bandwidth of 10 points and a linear polynomial specification.

Our identification strategy is perhaps most easily understood by observing a sequence of figures. Online Appendix Figure 5 presents the building blocks of a DiD in RD design in which the dependent variable is the probability of remediation under the discretion versus rule regime for different groups of students. Panels A and B present RD graphs showing differences in remediation probabilities for students near the proficiency threshold in the discretion and rule regimes, respectively, and Panel C presents the difference in the RD between the discretion

versus the rule regime. This is a zoomed-in version of Panels A and C in Figure 1, and it is immediately clear that students just below the proficiency threshold were dramatically less likely to be remediated under discretion than under the rule regime. Our identification strategy, however, relies on there being different levels of treatment intensity for marginal students across regimes. Panels D-F repeat these same comparisons for students whom we predict to be relatively highly likely to be remediated under discretion, while Panels G-I present similar comparisons for students whom we predict to be relatively unlikely to be remediated under discretion. Our DiD in RD estimate is the difference between the discontinuities observed in Panel F and Panel I, and the DDD estimate is its counterpart. To further fix ideas, Online Appendix Figure 6 presents similar graphs for one of our dependent variables of interest – postsecondary degree attainment.

5. Results

5.1 Evidence that education discretion may improve student outcomes relative to rules

Table 3 and Table 4 present the results from the DDD and DiD in RD analysis, looking at the effects on postsecondary enrollment and persistence in Table 3, and the effects on degree attainment in Table 4. In each table, the first three columns present the means (and standard deviations in parentheses) of the dependent variable for never-flagged students within 10 points below the cutoff under the discretion regime (first column), the rule regime (second column), and the estimated difference between discretion and rule regimes (third column) for students with low predicted likelihood and high predicted likelihood of remediation (first and second rows of each panel). Columns four through six repeat the same analysis for students within 10 points above the cutoff. The first two rows of the last column in each panel present the DiD estimates

taking the difference in differences presented in columns three and six. Finally, the last two rows of each panel present the DDD estimates (difference in the DiD estimates for the low predicted likelihood versus high predicted likelihood students) and DiD in RD estimates. Figure 3 and Figure 4 depict the DiD in RD analysis graphically, presenting the difference in raw cell means of the dependent variable between low- and high-predicted likelihood students under the discretion regime and the rule regime for postsecondary enrollment and persistence in Figure 3 and degree attainment in Figure 4.

The findings suggest significant positive effects of discretion on student outcomes. For example, DDD estimates suggest that discretion improves postsecondary enrollment by 9 percentage points (roughly equivalent to 20 percent of the dependent variable mean for students in the analytic sample), persistence in college beyond second year by 11 percentage points (about 40 percent of the dependent variable mean), 2-year degree attainment by 10 percentage points (about 50 percent of the dependent variable mean), and 4-year degree attainment by 4 percentage points although this estimate is marginally significant at conventional levels. DiD in RD results also point in the same direction, but they are smaller in magnitude and not statistically significant for postsecondary enrollment and 4-year degree attainment (as also indicated by the smaller positive discontinuities for these outcomes under the discretion regime in Figure 1 and Figure 2).

5.2. Effects on intermediate outcomes and educational inputs

Table 5 presents the DDD and DiD in RD results for intermediate outcomes (and educational inputs) that might explain the effects on postsecondary outcomes including the likelihood of being placed in remedial schedule in ELA, classroom-peer characteristics in ELA and other core courses (math, science, and social studies), and teacher value-added in ELA

during the following school year as well as overall high school outcomes such as advanced course-taking in high school (number of advanced courses and AP courses taken, likelihood of passing an AP test, number of AP tests passed) and high school graduation.

The DDD estimates suggest that the discretion regime has significant effects on both educational inputs during the following year (the year of potential remediation) and overall high school outcomes. In particular, we find that the discretion regime significantly reduces the likelihood that never-flagged students are placed in the remedial schedule in ELA; significantly increases average classroom-peer prior achievement not only in the remediation subject (by 27 percent of a standard deviation [0.27σ]), but also in other core subjects by 0.13σ; reduces the share of subsidized meal eligible peers by 8 percentage points in ELA and 6 percentage points in other core subjects, and increases teacher value-added scores in ELA courses by 0.01σ in student ELA scores (about a third of the standard deviation in teacher value-added scores).

Similarly, the discretion regime also improve the advanced course-taking and passing rates in high school, increasing the number of advanced courses taken in high school by nearly one (about 12 percent of the dependent variable mean), number of AP courses taken by 0.7 (20 percent of the dependent variable mean), likelihood of passing an AP test by 6 percentage points (30 percent), number of AP tests passed by 0.17 (10 percent). We find no significant effect on high school graduation. RD in DD estimates mostly align with the DDD results, but they are imprecisely estimated.

To what extent can the postsecondary effects of discretion be explained by the effect on high school outcomes? Table 5A takes a closer look at this question and examines whether certain high school outcomes serve as stronger mediators behind the observed effects on postsecondary outcomes. Column (I) of this table presents the DDD estimates on the

corresponding outcome of interest (postsecondary enrollment in the top panel; persistence beyond second year in the second panel; 2-year degree attainment in the third panel; and 4-year degree attainment in the bottom panel) reported in Table 3 and Table 4 whereas column (II) adds the high school outcome listed in the first column as a covariate.

The findings suggest that the effect of the discretion regime on classroom-peer characteristics in high school explains a considerable portion of the total effect of remediation on postsecondary outcomes. For example, the positive effect of the discretion regime on classroom-peer characteristics explains roughly 65 percent of its effect on postsecondary enrollment, about half of its effect on persistence in college; 30 percent of its effect on 2-year degree attainment, and 75 percent of its effect on 4-year degree attainment. The results also suggest that advanced course-taking in high school is an important mediator behind the observed postsecondary effects, while the effects on high school graduation fail to explain much of its effects on postsecondary enrollment and graduation.

5.3. Heterogeneity

Table 6 breaks down the DDD analysis on postsecondary outcomes in Table 2 and Table 3 by student attributes including race/ethnicity, subsidized meal eligibility, language spoken at home, and gender. Aside from the postsecondary outcomes, the last column presents results for remedial placement in the following year to examine whether discretion has different implications on remediation for different student groups. Similarly, we present school-level heterogeneity results in Table 7, breaking down the analysis by school racial composition, school poverty, and school test performance (mean test scores, variation in test scores, and the share of students flagged for remediation). There are several findings worth highlighting.

First, we find weak evidence of differential treatment under discretion: The last column of Table 6 suggests large reduction in remediation rates for low-probability students across all groups, but reduction is largest for Hispanic students, students who were always categorized as eligible for subsidized meals, English non-native students, and male students. Second, we show that this reduction has different implications for postsecondary outcomes for different student groups while discretion treatment effects are always positive (albeit not always statistically significant) for all subgroups. For example, we find that 2-year degree attainment effects are strongest for Hispanic students, English non-native students, males, yet they are similar for affluent vs. less-affluent students. In contrast, 4-year degree attainment results are stronger for White students, English native students, and relatively affluent students. Similarly, the findings presented in Table 7 suggest that the 2-year degree attainment effects are larger for schools that serve higher shares of non-White students, and higher shares of subsidized meal eligible students while the 4-year degree attainment effects are somewhat larger for other schools (although the differences in estimated effects are not statistically different from zero at conventional levels).

5.4. Is discretion just "less of a bad treatment"?

Are these effects driven by some students avoiding a "bad treatment" under the discretion regime or are the positive effects driven by educators better identifying students who would benefit more from the treatment under the discretion regime? To answer this question, we conduct two exercises in cases where some students experienced a positive treatment shock under different regimes, and we examine how their high school and postsecondary outcomes changed (compared to observationally similar students). The idea here is that if it is indeed

"avoiding a bad treatment" that drives our previous results, we would expect to find negative effects of these treatment shocks.

First, we leverage the difference in treatment likelihood that never-flagged, high probability students face under discretion and no test regimes when they score right below the remediation cutoff. As shown in Figure 1, never-flagged students were almost never remediated under the no test regime even if they scored below the cutoff while 61 percent of these high-probability students were remediated under the discretion regime (shown in the top panel of Table 2). Using this variation in treatment likelihood, we conduct a DDD exercise similar to our earlier analysis, comparing high probability versus low remediation probability students, under discretion versus no test regimes, and for those right below versus right above the remediation cutoff.[11]

Tables 8 through 12 present our findings for postsecondary enrollment and persistence (Table 8), degree attainment (Table 9), intermediate high school outcomes (Table 10), heterogeneity results based on student attributes (Table 11), and heterogeneity results based on school attributes (Table 12). Overall, we find precisely estimated zero effects overall and for different subgroups even though discretion regime increases remediation likelihood for students in the treatment group by 39 percentage points (as shown in the first row of Table 10). These findings present evidence against the hypothesis that the earlier positive effects are driven by students avoiding a "bad treatment".

Second, we leverage the difference in treatment likelihood students who score right above the remediation cutoff face under the no test and rule regimes based on whether they were

[11] Similar to the discretion versus rule analysis, we obtain predicted probabilities using never-flagged students within 10 points below the remediation cutoff. Because we have fewer cohorts in this analysis (as we have 3 school years instead of 5), we define low- and high-probability students based on the median of the distribution (rather than terciles) to preserve statistical power.

flagged for remediation in the previous year. As shown in Figures 1 and 2, under the no test regime, many students right above the cutoff were remediated if they were flagged for remediation in the previous year (Figure 2 illustrates this using previously remediated students) while very few students right above the cutoff were remediated if they were not flagged in the previous year (as shown in Figure 1). In contrast, students were rarely remediated under the rule regime if they scored above the cutoff. Using this variation, we conduct a DiD exercise comparing students within 10 points above the cutoff that were flagged for remediation in the previous year with those that were not flagged, under the no test regime versus the discretion regime. We conduct this exercise separately based on students' predicted remediation probability under the discretion regime[12]: The idea here is that high-probability students might benefit more from the treatment under the no test regime compared to low-probability students if educators indeed identify students who would benefit more from treatment under the discretion regime.

Tables 13 through 17 present the results with the remediation placement and postsecondary enrollment results provided in Table 13, degree attainment provided in Table 14, intermediate outcomes in Table 15, heterogeneity analysis based on student attributes in Table 16, and based on school attributes in Table 17. The findings suggest that the no test regime significant increases remediation rates for both high-probability and low-probability students (DiD estimate of 69 percentage points for high-probability students versus 59 percentage points for low-probability students) if they were flagged for remediation in the previous year. That said, the postsecondary consequences of increased remediation for these student groups are

[12] We obtain predicted probabilities using students within 10 points below the remediation cutoff that were not flagged for remediation in the previous year. Because we have fewer cohorts in this analysis (as we have 4 school years instead of 5), we define low- and high-probability students based on the median of the distribution (rather than terciles) to preserve statistical power.

considerably different. We find positive (and statistically significant) effects on postsecondary enrollment, persistence, and 4-year degree attainment for high-probability students while we find zero (and in some cases negative, yet statistically insignificant) effects for low-probability students, leading to positive and statistically significant consequences of increased remediation for these outcomes. The heterogeneity analysis reveal that the postsecondary enrollment and persistence effects are somewhat more pronounced for students from more disadvantaged backgrounds (e.g., students who were always classified as eligible for subsidized meals) and for schools serving more disadvantaged student populations while the 4-year degree attainment effects are more pronounced for more advantaged students and schools.

6. Conclusion

Our analysis of Florida's shift from a generally strongly-enforced rule-based mechanism for determining which middle and high school students were remediated to one where educators had considerably more discretion to make remediation decisions provides an excellent opportunity to gauge whether providing professional discretion in an environment in which professionals have a tremendous amount of private information about students' capacity and needs may be beneficial. We suspect that this large degree of private information may be an explanation for why affording educators professional discretion apparently leads to improved long-run student outcomes, while professionals with little private information (e.g., the hiring managers who chose to ignore testing results in Hoffman, Kahn, and Li's (2018) study) apparently tend to make worse decisions when exercising discretion. Physicians prescribing medications off-label (in Cuddy and Currie's (forthcoming) case) likely have more valuable private information than the group whom Hoffman, Kahn, and Li study, but they don't have the

daily interactions with students, together with frequent assessments, over the course of nine months of schooling that educators have. Understanding the role of relatively extensive private information in effective deployment of professional discretion seems like a highly fruitful area for future research.

On the policy and practice topic at hand: Our results, taken together, provide evidence that some degree of remediation is valuable, but that it is most likely to be valuable when exercised judiciously by educators relying upon a wide range of information – including test scores. The test score piece seems important, because we find evidence that proficiency thresholds seem to help shine a light on the students who may need more help. When educators had discretion but no new test score information, they likely under-remediated students who could have benefited from remediation. On the flip side, the lack of a bright line in 2015 ended up facilitating beneficial continued remediation for some students who would surely not have received this help if the bright line were visible. (It is unknown the degree to which this lack of continued remediation when bright lines are visible is due to educator decision-making versus parental preferences, but this is another important question to assess in future work on the subject.) One major challenge for educators may be to use proficiency thresholds as a place to focus attention – but not make these thresholds dispositive for students on *either* the "proficient" or "not proficient" side of the line.

References

Borcan, O., Lindahl, M., & Mitrut, A. (2017). Fighting corruption in education: What works and who benefits? *American Economic Journal: Economic Policy* 9(1): 180-209.

Busso, M., Montaño, S., Muñoz-Morales, J., & Pope, N. (2024). The unintended consequences of merit-based teacher selection: Evidence from large-scale reform in Colombia. NBER working paper 33008.

Carril, R. (2022). Rules versus discretion in public procurement. Working paper, Universitat Pompeu Fabra.

Cuddy, E. & Currie, J. (forthcoming). Rules vs. discretion: Treatment of mental illness in U.S. adolescents, *Journal of Political Economy*, forthcoming.

Dee, T. S., Dobbie, W., Jacob, B., & Rockoff, J. E. (2019). The causes and consequences of test score manipulation: Evidence from the New York Regents Examinations. *American Economic Journal: Applied Economics, 11*(3), 382–423.

Diamond, R., & Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes exams. NBER working paper 22207.

Eren, O., Depew, B., & Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics*, 152: 9–31.

Eren, O., Lovenheim, M., & Mocan, N. (2022). The effect of grade retention on adult crime: Evidence from a test-based promotion policy. *Journal of Labor Economics*, 40(2): 361–395.

Figlio, D., & Özek, U. (2020). An extra year to learn English? Early grade retention and the human capital development of English learners. *Journal of Public Economics, 186*, 104158.

Figlio, D., & Özek, U. (2024). The unintended consequences of test-based remediation. *American Economic Journal: Applied Economics*, 16(1): 60-89.

Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy, 2*(4), 319–340.

Greene, J. P., & Winters, M. A. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy, 7*(3), 305–330.

Hoffman, M., Kahn, L. B., & Li, D. (2018). Discretion in hiring. *Quarterly Journal of Economics, 133*(2), 765–800.

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

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

Jacob, B. & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics*, *1(3)*: 33–58.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics, 133*(1), 237–293.

Larsen, M. & Valant, J. (2018). The long-term effects of grade retention: Evidence on persistence through high school and college. Unpublished manuscript.

Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis*, 35(1), 96-117.

Mariano, L., Martorell, P., & Berglund, T. (2018). The effects of grade retention on high school outcomes: Evidence from New York City schools. Working paper.

Mumma, K., & Winters, M. (2023). The effect of retention under Mississippi's test-based promotion policy. Wheelock Educational Policy Center working paper 2023-1.

Oh, S. (2025). Does discretion by elected judges benefit defendants, or judges? Evidence from public defense case assignment. Working paper, Texas A&M University.

Özek, U. (2021). The Effects of Middle School Remediation on Postsecondary Success: Regression Discontinuity Evidence from Florida. *Journal of Public Economics*, 203, 104518.

Özek, U. (2025). The Effects of High School Remediation on Long-Run Educational Attainment *Journal of Human Resources*, forthcoming.

Papay, J. P., Murnane, R. J., & Willett, J. B. (2014). High-school exit examinations and the schooling decisions of teenagers: Evidence from regression-discontinuity approaches, *Journal of Research on Educational Effectiveness* 7(1): 1-27.

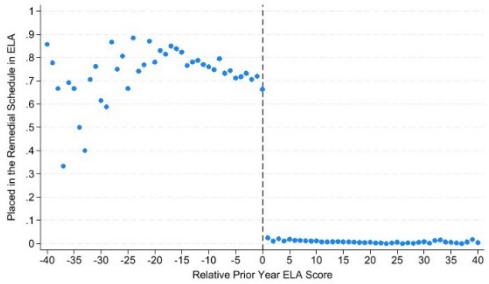
Ricks, Michael (2024). Self-selection around policy recommendations: The case of kindergarten entry, University of Nebraska working paper.

Rockoff, Jonah, Thomas Kane, Douglas Staiger, and Eric Taylor (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools, *American Economic Review* 102(7): 3184-3213.

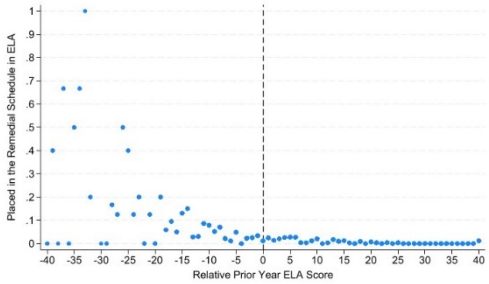
Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics, 152*, 154–169.

Figure 1: Likelihood of Being Placed in Remedial ELA Schedule under the Three Regimes, Never Flagged for Remediation Previously

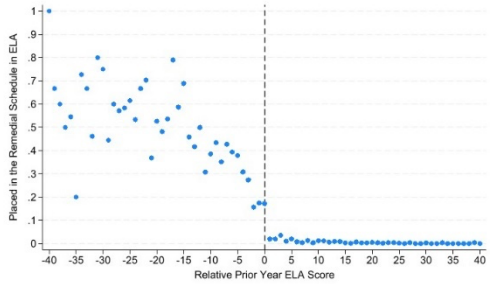
(A) Rule Regime



(B) No Test Regime



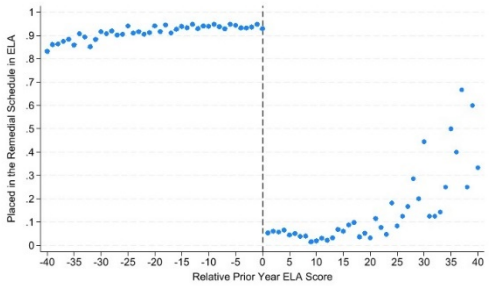
(C) Discretion Regime



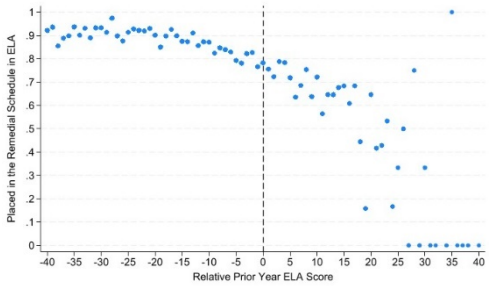
Note: Each figure presents the likelihood of being placed in the remedial ELA schedule in grades 9 and 10, averaged at the prior year ELA scale score level between 40 points below and 40 points above the remediation cutoff during rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18) for students who never scored below the remediation cutoff in the prior two years.

Figure 2: Likelihood of Being Placed in Remedial ELA Schedule under the Three Regimes, Remediated During the Test Year

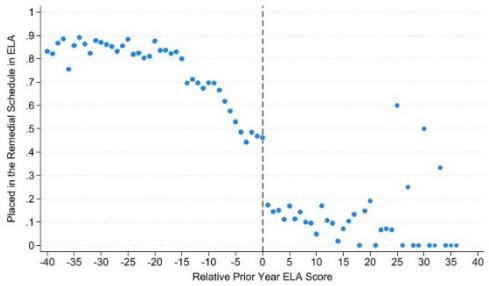
(A) Rule Regime



(B) No Test Regime



(C) Discretion Regime

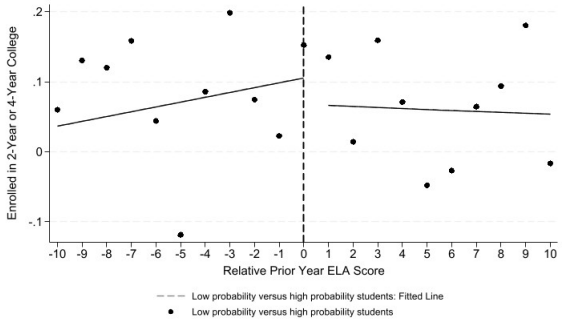


Note: Each figure presents the likelihood of being placed in the remedial ELA schedule in grades 9 and 10, averaged at the prior year ELA scale score level between 40 points below and 40 points above the remediation cutoff during rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18) for students who were in the remedial schedule during the test year.

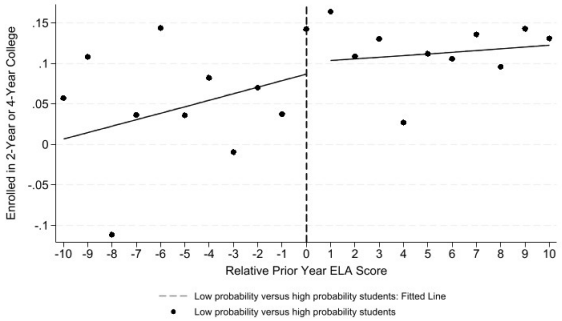
Figure 3. Postsecondary Enrollment and Persistence Around the Remediation Cutoff, by Regime and Predicted Probability of Remediation, Never Flagged for Remediation, Discretion versus Rule Regime

Enrolled in a 2- or 4-Year College

(A) Discretion Regime

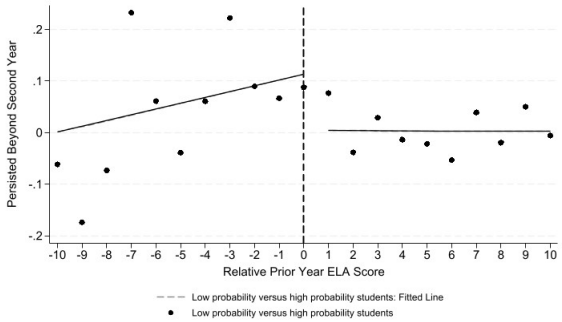


(B) Rule Regime

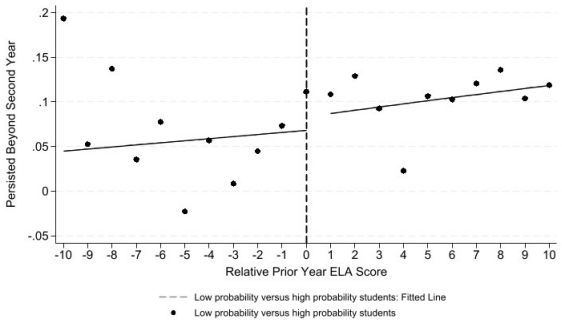


Persisted Beyond Second Year in College

(C) Discretion Regime



(D) Rule Regime

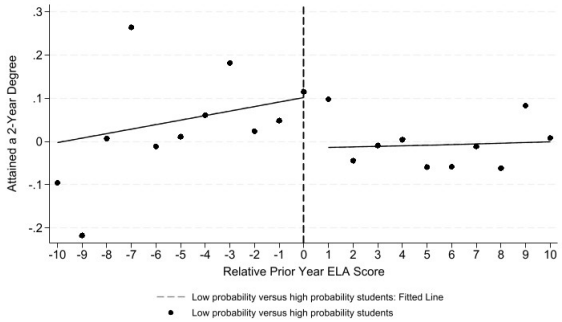


Notes: Each figure presents the difference in raw cell means of postsecondary enrollment (top panel) and persistence beyond second year in college (bottom panel) for each prior year ELA score between 10 points below and 10 points above the remediation cutoff between low predicted probability and high predicted probability students (low ≤ 0.2 and high > 0.4) under the discretion regime (left panels) and rule regime (right panels). The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.

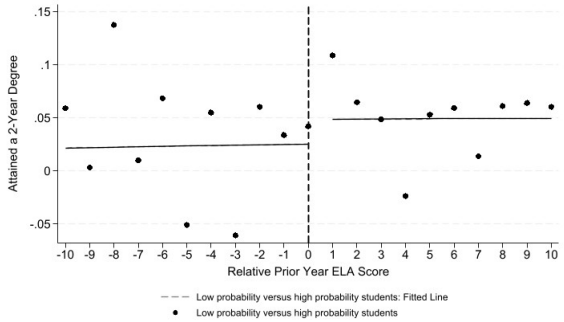
Figure 4. Postsecondary Degree Attainment Around the Remediation Cutoff, by Regime and Predicted Probability of Remediation, Never Flagged for Remediation, Discretion versus Rule Regime

Attained a 2-Year Degree

(A) Discretion Regime

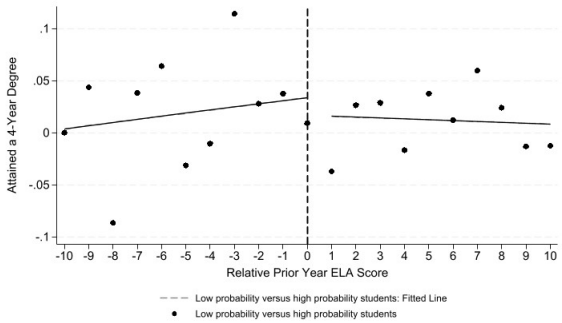


(B) Rule Regime

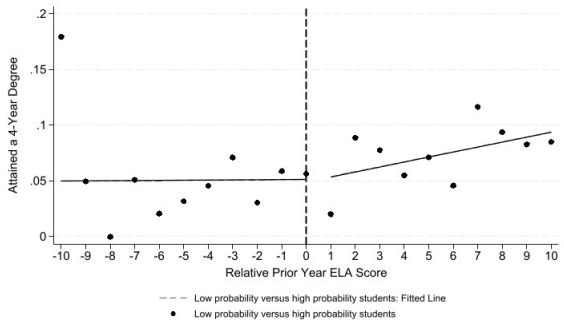


Attained a 4-Year Degree

(C) Discretion Regime

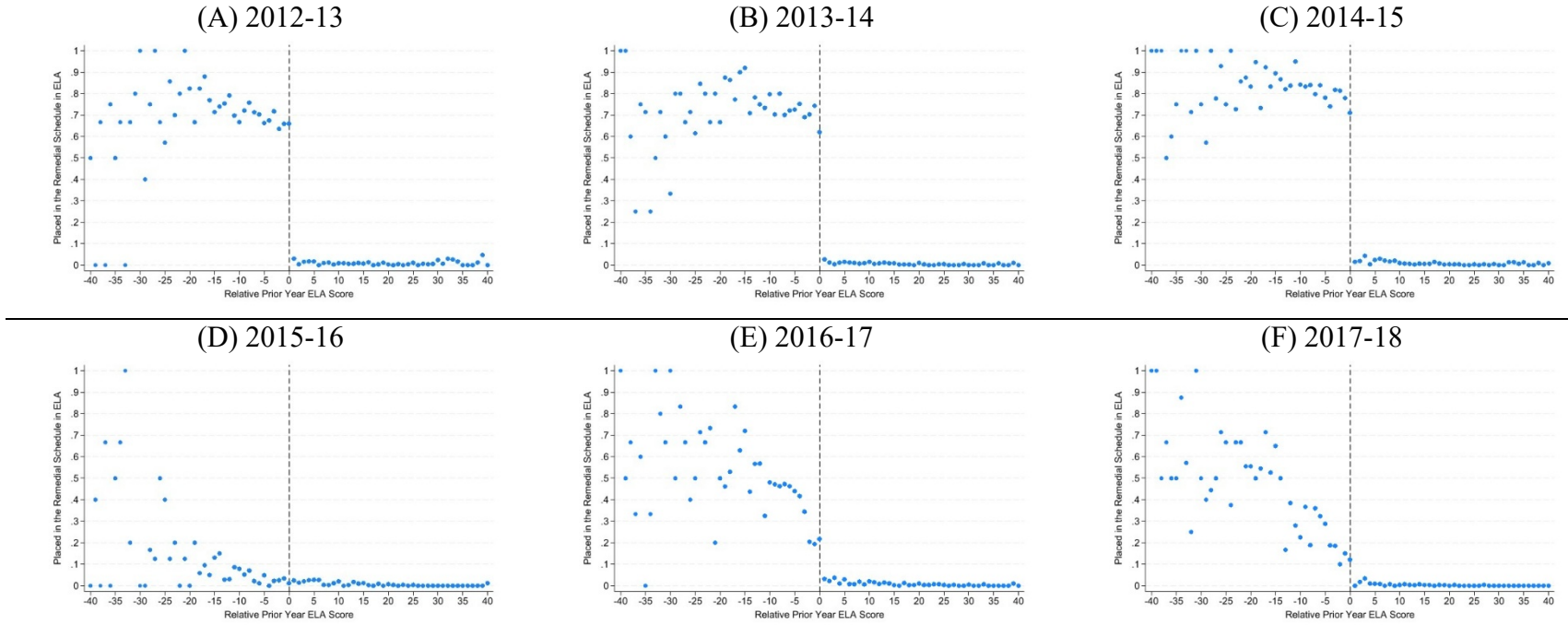


(D) Rule Regime



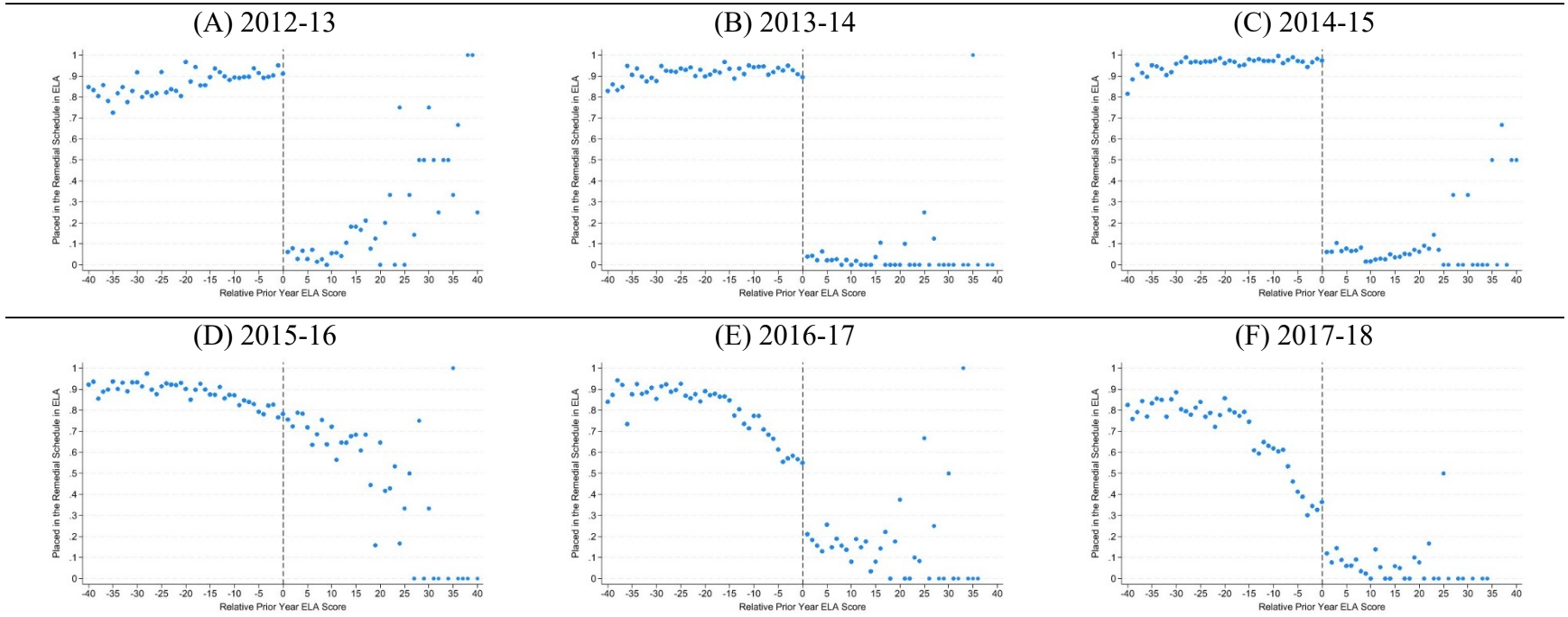
Notes: Each figure presents the difference in raw cell means of 2-year degree attainment (top panel) and 4-year degree attainment (bottom panel) for each prior year ELA score between 10 points below and 10 points above the remediation cutoff between low predicted probability and high predicted probability students (low ≤ 0.2 and high > 0.4) under the discretion regime (left panels) and rule regime (right panels). The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.

Online Appendix Figure 1: Likelihood of Being Placed in Remedial ELA Schedule under the Three Regimes, Never Flagged for Remediation Previously



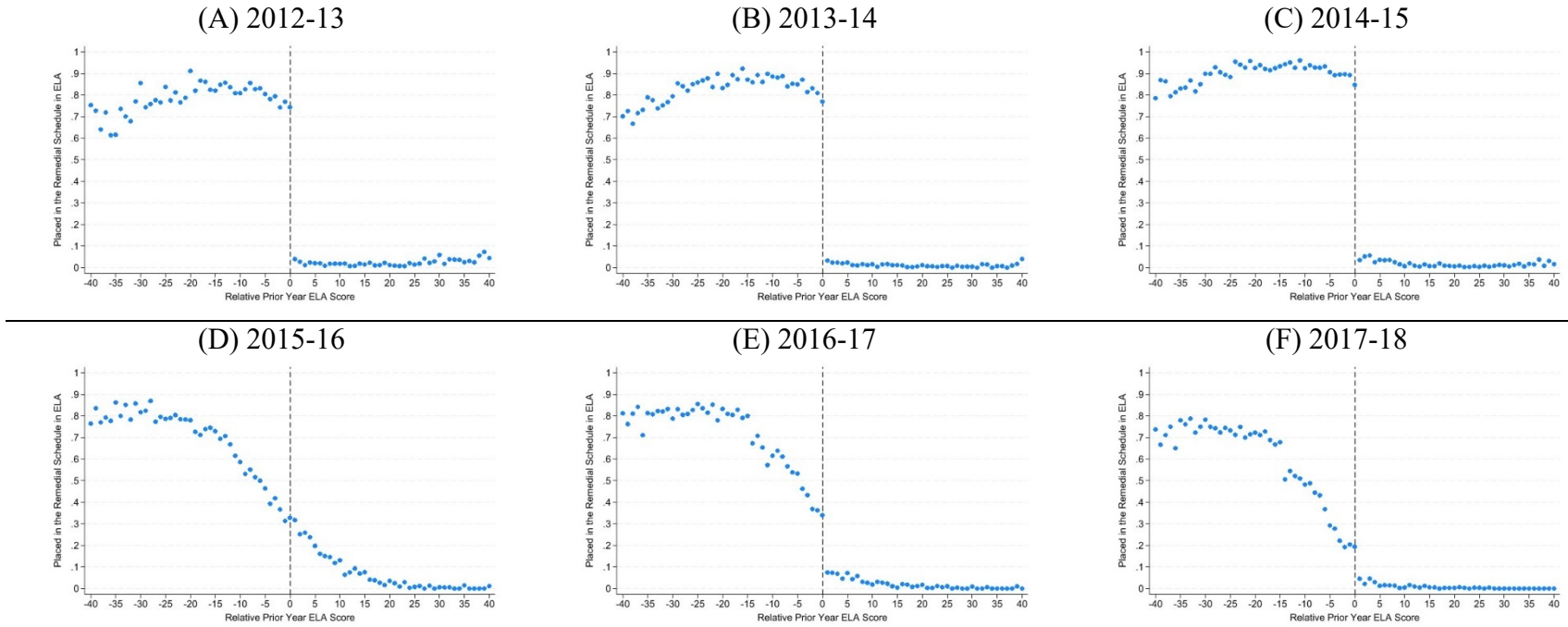
Note: Each panel presents the likelihood of being placed in the remedial ELA schedule in grades 9 and 10, averaged at the prior year ELA scale score level between 40 points below and 40 points above the remediation cutoff during rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18) for students who were never flagged for remediation in the prior two years.

Online Appendix Figure 2: Likelihood of Being Placed in Remedial ELA Schedule under the Three Regimes, Remediated During the Test Year



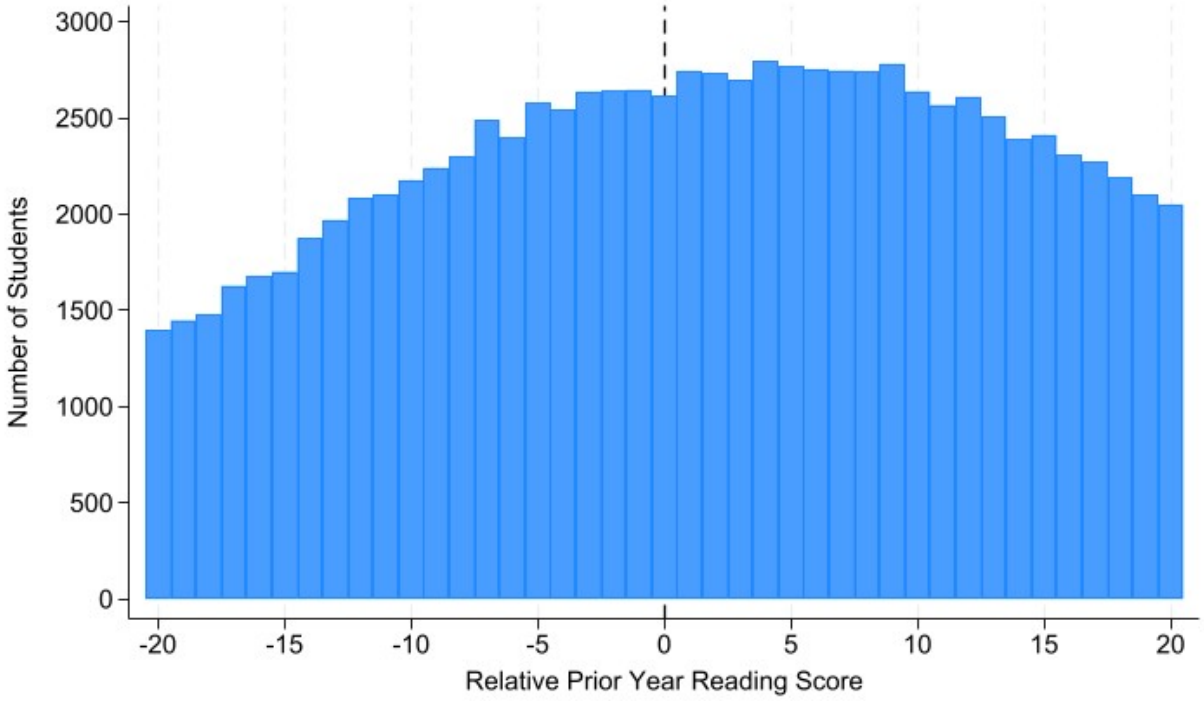
Note: Each panel presents the likelihood of being placed in the remedial ELA schedule in grades 9 and 10, averaged at the prior year ELA scale score level between 40 points below and 40 points above the remediation cutoff during rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18) for students who were placed in the remedial schedule during the test year.

Online Appendix Figure 3: Likelihood of Being Placed in Remedial ELA Schedule under the Three Regimes, Overall



Note: Each panel presents the likelihood of being placed in the remedial ELA schedule in grades 9 and 10, averaged at the prior year ELA scale score level between 40 points below and 40 points above the remediation cutoff during the rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18).

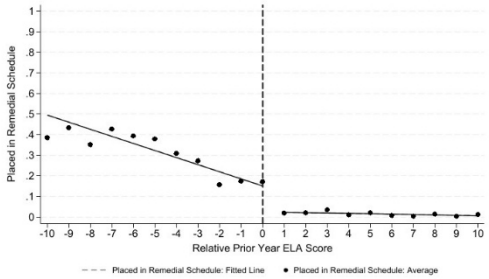
Online Appendix Figure 4. Prior Year Reading Score Density Around the Remediation Cutoff



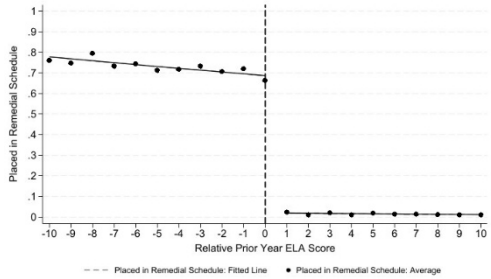
Notes: The figure presents the density of the running variable (prior year ELA score) centered at the remediation cutoff.

Online Appendix Figure 5: Remedial ELA Placement Around the Remediation Cutoff, by Regime and Predicted Probability of Remediation, Never Flagged for Remediation, Discretion versus Rule Regime

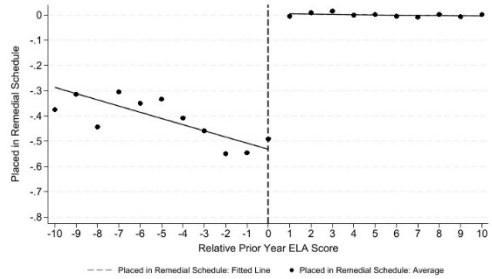
(A) Discretion Regime



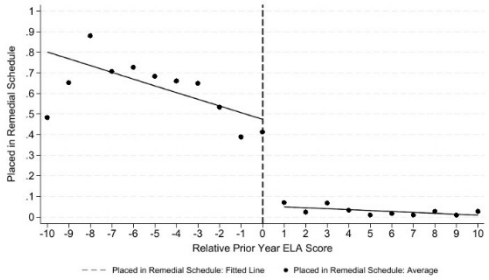
(B) Rule Regime



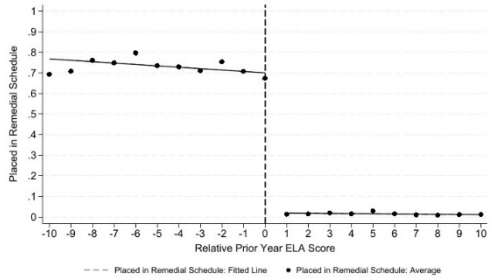
(C) Discretion versus Rule Regime



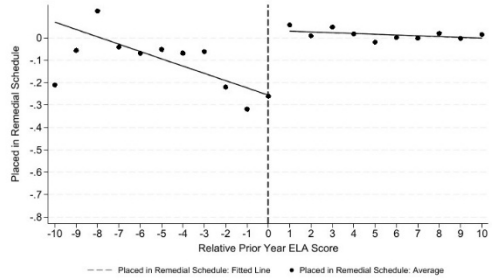
(D) Discretion Regime – High Likelihood of Remediation



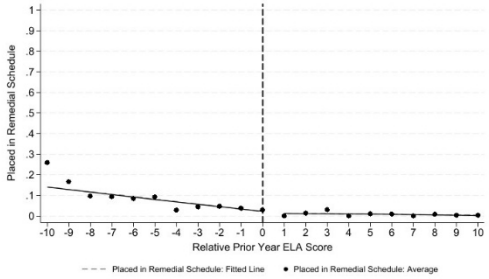
(E) Rule Regime – High Likelihood of Remediation



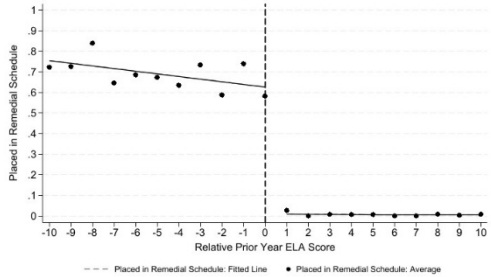
(F) Discretion versus Rule Regime – High Likelihood of Remediation



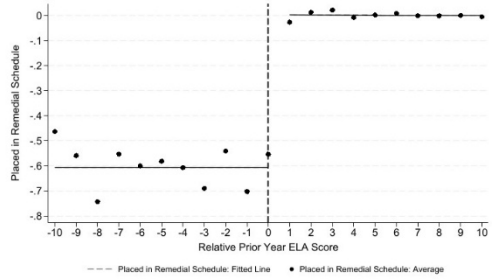
(G) Discretion Regime – Low Likelihood of Remediation



(H) Rule Regime – Low Likelihood of Remediation



(I) Discretion versus Rule Regime – Low Likelihood of Remediation

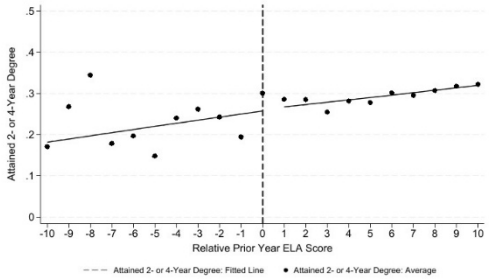


Notes: Each figure presents the raw cell means of 2- or 4-year degree attainment for each prior year ELA score between 10 points below and 10 points above the remediation cutoff under discretion regime (left panel), rule regime (middle panel), discretion versus rule regime (right panel) overall (first row), for high

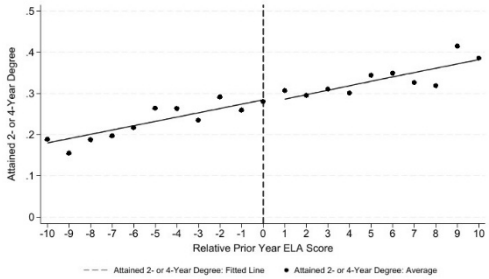
predicted probability students (second row), and for low predicted probability students (bottom row, low ≤ 0.2 and high > 0.4). The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.

Online Appendix Figure 6: Postsecondary Degree Attainment Around the Remediation Cutoff, by Regime and Predicted Probability of Remediation, Never Flagged for Remediation, Discretion versus Rule Regime

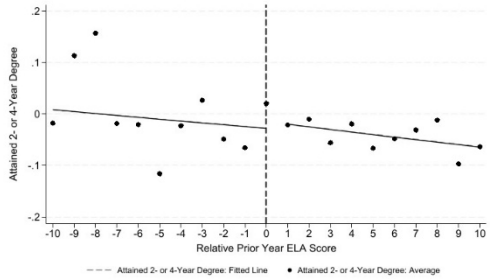
(A) Discretion Regime



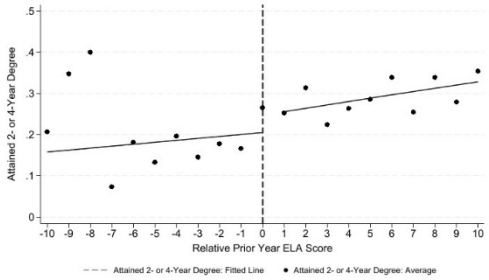
(B) Rule Regime



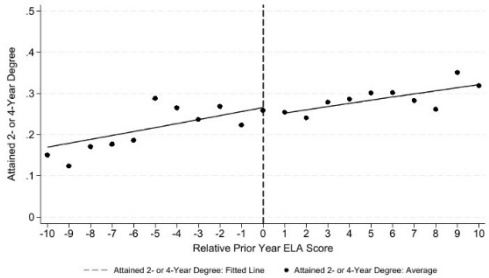
(C) Discretion versus Rule Regime



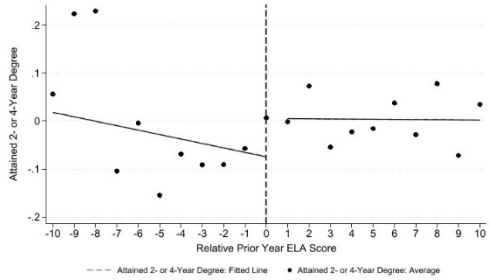
(D) Discretion Regime – High Likelihood of Remediation



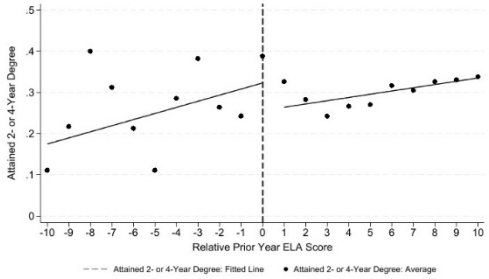
(E) Rule Regime – High Likelihood of Remediation



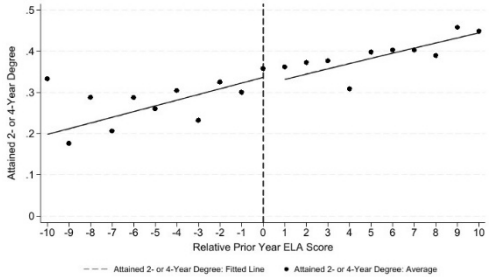
(F) Discretion versus Rule Regime – High Likelihood of Remediation



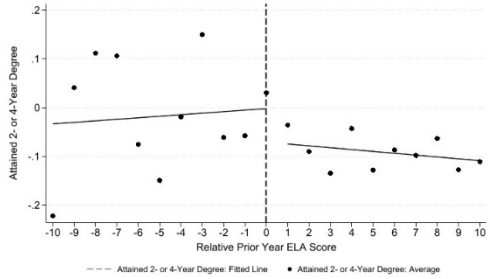
(G) Discretion Regime – Low Likelihood of Remediation



(H) Rule Regime – Low Likelihood of Remediation



(I) Discretion versus Rule Regime – Low Likelihood of Remediation



Notes: Each figure presents the raw cell means of 2- or 4-year degree attainment for each prior year ELA score between 10 points below and 10 points above the remediation cutoff under discretion regime (left panel), rule regime (middle panel), discretion versus rule regime (right panel) overall (first row), for high

predicted probability students (second row), and for low predicted probability students (bottom row, low ≤ 0.2 and high > 0.4). The black solid line and the dashed gray lines provide the fitted lines estimated separately for observations below and above the cutoff, which is shown by the vertical line. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.

Table 1: Student Attributes Around the Remediation Cutoff under the Three Regimes, Students 0-10 Points Above versus Below

	Overall		Rule Regime		No Test Regime		Discretion Regime	
	Above	Below	Above	Below	Above	Below	Above	Below
White	0.310	0.236	0.315	0.239	0.318	0.238	0.300	0.230
	(0.463)	(0.425)	(0.465)	(0.427)	(0.466)	(0.426)	(0.458)	(0.421)
Black	0.247	0.311	0.250	0.323	0.238	0.306	0.247	0.295
	(0.431)	(0.463)	(0.433)	(0.468)	(0.426)	(0.461)	(0.432)	(0.456)
Hispanic	0.364	0.392	0.357	0.377	0.366	0.391	0.376	0.417
	(0.481)	(0.488)	(0.479)	(0.485)	(0.482)	(0.488)	(0.484)	(0.493)
Male	0.488	0.513	0.495	0.509	0.483	0.511	0.480	0.520
	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)	(0.500)
English non-native	0.301	0.329	0.287	0.319	0.304	0.321	0.321	0.350
	(0.459)	(0.470)	(0.452)	(0.466)	(0.460)	(0.467)	(0.467)	(0.477)
Subsidized meal eligible	0.602	0.698	0.583	0.681	0.593	0.699	0.636	0.725
	(0.489)	(0.459)	(0.493)	(0.466)	(0.491)	(0.459)	(0.481)	(0.447)
Special education – prior year	0.057	0.0896	0.071	0.103	0.051	0.079	0.039	0.073
	(0.232)	(0.286)	(0.257)	(0.304)	(0.220)	(0.270)	(0.193)	(0.261)
English learner – prior year	0.029	0.0593	0.025	0.060	0.032	0.053	0.034	0.062
	(0.167)	(0.236)	(0.155)	(0.237)	(0.175)	(0.224)	(0.181)	(0.240)
ELA score – two years prior	0.165	-0.225	0.133	-0.254	0.173	-0.211	0.212	-0.184
	(0.511)	(0.506)	(0.516)	(0.503)	(0.527)	(0.509)	(0.491)	(0.509)
High performer in math	0.481	0.278	0.472	0.269	0.476	0.274	0.479	0.278
	(0.500)	(0.448)	(0.499)	(0.444)	(0.499)	(0.446)	(0.500)	(0.448)
Disciplinary incident – prior year	0.195	0.275	0.246	0.338	0.142	0.214	0.143	0.201
	(0.396)	(0.446)	(0.431)	(0.473)	(0.349)	(0.410)	(0.350)	(0.401)
% absent days – prior year	0.047	0.0541	0.047	0.055	0.043	0.049	0.049	0.056
	(0.049)	(0.0547)	(0.051)	(0.058)	(0.043)	(0.047)	(0.049)	(0.053)
Remedial ELA course – prior year	0.193	0.411	0.187	0.406	0.234	0.491	0.180	0.377
	(0.394)	(0.492)	(0.390)	(0.491)	(0.423)	(0.500)	(0.384)	(0.485)
Remedial math course – prior year	0.160	0.266	0.142	0.246	0.213	0.346	0.159	0.258
	(0.366)	(0.442)	(0.349)	(0.431)	(0.410)	(0.476)	(0.366)	(0.437)
N	40,192	37,733	13,947	13,095	4,476	4,045	8,973	7,947

Notes: Standard deviations are given in parentheses. Test scores are standardized at the grade-year level to zero mean and unit variance. The first two columns present summary statistics for all students within 10 points around the remediation cutoff between 2012-13 and 2017-18 school years whereas the other columns compare these students under the rule regime (2012-13 through 2014-15), no test regime (2015-16), and discretion regime (2016-17 and 2017-18). High performer in math is defined as students with above-median average math scores on prior year tests.

Table 2: Remediation Probabilities for the Three Regimes by Predicted Remediation Likelihoods under the Discretion Regime, Never-flagged Students 0-10 points Below and Above the Cutoff

	0-10 points below the cutoff		
	Low predicted likelihood	Medium predicted likelihood	High predicted likelihood
Rule regime	0.670	0.769	0.728
	(0.471)	(0.421)	(0.445)
No test regime	0.012	0.014	0.032
	(0.107)	(0.117)	(0.176)
Discretion regime	0.056	0.256	0.613
	(0.230)	(0.437)	(0.488)
	0-10 points above the cutoff		
	Low predicted likelihood	Medium predicted likelihood	High predicted likelihood
Rule regime	0.006	0.012	0.015
	(0.079)	(0.107)	(0.121)
No test regime	0.004	0.022	0.025
	(0.059)	(0.147)	(0.156)
Discretion regime	0.008	0.010	0.028
	(0.086)	(0.100)	(0.165)

Note: Each panel presents the average remediation probability by predicted remediation likelihood (low ≤ 0.2; $0.2 <$ medium ≤ 0.4; high > 0.4) and placement regime conditional on having never been flagged for remediation in the previous two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.

Table 3: Discretion versus Rules: Effects on Postsecondary Enrollment and Persistence, Difference-in-Difference-in-Differences and Differences-in-Differences in Regression Discontinuity Results

	Postsecondary enrollment						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	Rule regime	Difference	Discretion regime	Rule regime	Difference	Difference-in-differences
Low predicted likelihood	0.497	0.537	-0.040	0.575	0.650	-0.076***	0.036
	(0.500)	(0.499)	(0.025)	(0.495)	(0.477)	(0.014)	(0.029)
High predicted likelihood	0.406	0.474	-0.068***	0.513	0.532	-0.019	-0.049
	(0.492)	(0.499)	(0.026)	(0.500)	(0.499)	(0.018)	(0.032)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.085**
							(0.043)
Difference-in-differences in regression discontinuity: low versus high predicted likelihood							0.044
							(0.081)
	Persisted beyond second year						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	Rule regime	Difference	Discretion regime	Rule regime	Difference	Difference-in-differences
Low predicted likelihood	0.264	0.287	-0.023	0.283	0.390	-0.107***	0.084***
	(0.441)	(0.453)	(0.023)	(0.451)	(0.488)	(0.014)	(0.026)
High predicted likelihood	0.192	0.225	-0.033	0.278	0.282	-0.004	-0.029
	(0.394)	(0.418)	(0.021)	(0.448)	(0.450)	(0.016)	(0.027)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.113***
							(0.037)
Difference-in-differences in regression discontinuity: low versus high predicted likelihood							0.124*
							(0.072)

Note: The two panels present the means (and standard deviations) of postsecondary enrollment (top panel) and persistence beyond second year in college (bottom panel) for students 0-10 points below and above the cutoff broken down by regime; differences between discretion and rule regimes for low and high predicted remediation likelihood groups (low ≤ 0.2 and high > 0.4); difference in these differences between students below and above the cutoff, and difference-in-difference-in-differences between low and high predicted likelihood students. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 4: Discretion versus Rules: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences and Differences-in-Differences in Regression Discontinuity Results

2-year degree attainment							
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	Rule regime	Difference	Discretion regime	Rule regime	Difference	Difference-in-differences
Low predicted likelihood	0.242	0.226	0.017	0.219	0.276	-0.057***	0.074***
	(0.429)	(0.418)	(0.022)	(0.413)	(0.447)	(0.013)	(0.025)
High predicted likelihood	0.176	0.197	-0.021	0.224	0.224	0.000	-0.021
	(0.381)	(0.398)	(0.020)	(0.417)	(0.417)	(0.015)	(0.025)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.095***
							(0.036)
Difference-in-differences in regression discontinuity: low versus high predicted likelihood							0.138**
							(0.068)
4-year degree attainment							
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	Rule regime	Difference	Discretion regime	Rule regime	Difference	Difference-in-differences
Low predicted likelihood	0.069	0.095	-0.026*	0.116	0.168	-0.051***	0.025
	(0.253)	(0.293)	(0.014)	(0.321)	(0.374)	(0.010)	(0.017)
High predicted likelihood	0.046	0.047	-0.001	0.103	0.0898	0.013	-0.015
	(0.209)	(0.212)	(0.011)	(0.304)	(0.286)	(0.011)	(0.015)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.040*
							(0.023)
Difference-in-differences in regression discontinuity: low versus high predicted likelihood							0.012
							(0.045)

Note: The two panels present the means (and standard deviations) of 2-year degree attainment (top panel) and 4-year degree attainment (bottom panel) for students 0-10 points below and above the cutoff broken down by regime; differences between discretion and rule regimes for low and high predicted remediation likelihood groups (low ≤ 0.2 and high > 0.4); difference in these differences between students below and above the cutoff, and difference-in-difference-in-differences between low and high predicted likelihood students. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 5: Discretion versus Rules: Effects on Intermediate Outcomes in High School

	DDD estimates	DD-in-RD estimates
Following year		
Placed in remedial schedule in ELA	-0.486***	-0.315***
	(0.031)	(0.055)
Classroom-peer prior test scores: ELA courses	0.265***	0.161**
	(0.038)	(0.073)
Classroom-peer prior test scores: other core courses	0.134***	0.041
	(0.038)	(0.072)
Classroom-peer subsidized meal eligible: ELA courses	-0.080***	-0.095***
	(0.019)	(0.037)
Classroom-peer subsidized meal eligible: other core courses	-0.059***	-0.065*
	(0.018)	(0.035)
Teacher value-added score in ELA courses	0.010***	0.004
	(0.003)	(0.006)
High school outcomes		
Number of advanced courses taken	0.977**	0.201
	(0.388)	(0.721)
Number of AP courses taken	0.650*	0.337
	(0.379)	(0.734)
Ever passed an AP test	0.060*	0.070
	(0.033)	(0.065)
Number of AP tests passed	0.165**	0.178
	(0.084)	(0.160)
Obtained a high school diploma	0.042	0.044
	(0.033)	(0.060)
Number of students	13,431	

Note: The first column presents the difference-in-difference-in-differences (DDD) estimates, comparing low versus high predicted remediation likelihood groups (low ≤ 0.2 and high > 0.4), below versus above the cutoff, and discretion versus rule regime. The DDD estimate provided represents the coefficient on the interaction between low remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The second column presents the difference-in-differences in regression discontinuity (DD in RD) estimates, comparing the discontinuity in the outcome of interest at the remediation cutoff for low versus high predicted remediation likelihood students under discretion versus rule regimes. DD in RD coefficients are obtained parametrically, using a linear polynomial and a bandwidth of 10 points. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 5A. Mediation Analysis: College Enrollment and Graduation Effects of High School Remediation

	Outcome: Enrolled in a 2-year or 4-year college	
Mediator	(I)	(II)
Peer characteristics in high school	0.085**	0.030
	(0.043)	(0.041)
Advanced course-taking in high school	0.085**	0.047
	(0.043)	(0.040)
High school diploma	0.085**	0.066
	(0.043)	(0.040)
	Outcome: Persisted beyond second year in college	
Mediator	(I)	(II)
Peer characteristics in high school	0.113***	0.064*
	(0.037)	(0.036)
Advanced course-taking in high school	0.113***	0.085**
	(0.037)	(0.036)
High school diploma	0.113***	0.103***
	(0.037)	(0.037)
	Outcome: Attained a 2-year degree	
Mediator	(I)	(II)
Peer characteristics in high school	0.095***	0.068*
	(0.036)	(0.035)
Advanced course-taking in high school	0.095***	0.082**
	(0.036)	(0.035)
High school diploma	0.095***	0.087**
	(0.036)	(0.035)
	Outcome: Attained a 4-year degree	
Mediator	(I)	(II)
Peer characteristics in high school	0.040*	0.010
	(0.023)	(0.022)
Advanced course-taking in high school	0.040*	0.023
	(0.023)	(0.022)
High school diploma	0.040*	0.037
	(0.023)	(0.023)
Mediator included	No	Yes

Notes: Column (I) of this table reports DDD estimates on postsecondary outcomes in Tables 3 and 4. Robust standard errors are given in parentheses. Column (II) introduces the high school outcomes listed in the first column as covariates. *, **, and *** represent statistical significance at 10, 5, and 1 percent, respectively.

Table 6: Discretion versus Rules: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by Student Attributes

		Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
White		0.171**	0.153**	0.060	0.116**	-0.462***
	N=4,342	(0.077)	(0.069)	(0.064)	(0.048)	(0.059)
Black		0.014	0.082	0.050	0.029	-0.406***
	N=3,222	(0.086)	(0.072)	(0.066)	(0.044)	(0.061)
Hispanic		0.093	0.162***	0.211***	0.004	-0.544***
	N=4,413	(0.072)	(0.060)	(0.059)	(0.030)	(0.051)
Always subsidized meal eligible		0.025	0.112**	0.139***	-0.005	-0.495***
	N=5,623	(0.063)	(0.053)	(0.050)	(0.027)	(0.043)
Never subsidized meal eligible		0.159*	0.209**	0.143*	0.126*	-0.422***
	N=3,590	(0.093)	(0.086)	(0.078)	(0.066)	(0.077)
English non-native		0.066	0.207***	0.264***	0.012	-0.556***
	N=3,622	(0.080)	(0.069)	(0.067)	(0.036)	(0.056)
English native		0.091*	0.075*	0.026	0.051*	-0.458***
	N=9,373	(0.051)	(0.045)	(0.042)	(0.029)	(0.038)
Male		0.113*	0.148***	0.130***	0.048*	-0.548***
	N=6,402	(0.060)	(0.049)	(0.047)	(0.028)	(0.043)
Female		0.056	0.078	0.062	0.030	-0.426***
	N=6,593	(0.061)	(0.056)	(0.053)	(0.036)	(0.045)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by student groups, comparing low versus high predicted remediation likelihood groups (low ≤ 0.2 and high > 0.4), below versus above the cutoff, and discretion versus rule regime. The DDD estimate provided represents the coefficient on the interaction between low remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 7: Discretion versus Rules: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by School Attributes

		Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
% White						
	Below median	0.086	0.109*	0.149***	0.018	-0.425***
	N=5,543	(0.065)	(0.056)	(0.054)	(0.031)	(0.046)
	Above median	0.074	0.091*	0.038	0.049	-0.546***
	N=6,914	(0.060)	(0.054)	(0.050)	(0.035)	(0.045)
% subsidized meal eligible						
	Below median	0.115*	0.122**	0.068	0.054	-0.416***
	N=6,503	(0.063)	(0.056)	(0.053)	(0.037)	(0.047)
	Above median	0.040	0.068	0.102**	0.010	-0.537***
	N=5,954	(0.063)	(0.053)	(0.050)	(0.030)	(0.043)
ELA scores – mean						
	Below median	0.062	0.054	0.059	0.037	-0.500***
	N=4,531	(0.072)	(0.056)	(0.053)	(0.030)	(0.051)
	Above median	0.088	0.116**	0.093*	0.031	-0.481***
	N=7,926	(0.056)	(0.052)	(0.049)	(0.033)	(0.041)
ELA scores – standard deviation						
	Below median	0.093	0.065	0.052	0.004	-0.514***
	N=6,837	(0.060)	(0.053)	(0.049)	(0.033)	(0.044)
	Above median	0.051	0.116**	0.116**	0.055	-0.431***
	N=5,620	(0.066)	(0.058)	(0.056)	(0.034)	(0.048)
% below remediation cutoff						
	Below median	0.113*	0.115**	0.090	0.030	-0.427***
	N=6,607	(0.064)	(0.058)	(0.055)	(0.038)	(0.048)
	Above median	0.031	0.061	0.066	0.028	-0.498***
	N=5,850	(0.063)	(0.052)	(0.050)	(0.029)	(0.044)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by school attributes, comparing low versus high predicted remediation likelihood groups (low ≤ 0.2 and high > 0.4), below versus above the cutoff, and discretion versus rule regime. The DDD estimate provided represents the coefficient on the interaction between low remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. School characteristics are calculated using student-level data between 2005-06 and 2011-

12 school years. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 8: Discretion versus No Test: Effects on Postsecondary Enrollment and Persistence, Difference-in-Difference-in-Differences Results

	Postsecondary enrollment						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	No test regime	Difference	Discretion regime	No test regime	Difference	Difference-in-differences
High predicted likelihood	0.423	0.457	-0.034	0.515	0.516	-0.000	-0.033
	(0.494)	(0.499)	(0.028)	(0.500)	(0.500)	(0.020)	(0.034)
Low predicted likelihood	0.497	0.547	-0.050*	0.562	0.594	-0.032**	-0.018
	(0.500)	(0.498)	(0.026)	(0.496)	(0.491)	(0.016)	(0.030)
		Difference-in-difference-in-differences: low versus high predicted likelihood					-0.016
							(0.045)
	Persisted beyond second year						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	No test regime	Difference	Discretion regime	No test regime	Difference	Difference-in-differences
High predicted likelihood	0.196	0.205	-0.009	0.272	0.301	-0.029	0.020
	(0.397)	(0.404)	(0.022)	(0.445)	(0.459)	(0.018)	(0.029)
Low predicted likelihood	0.255	0.301	-0.046*	0.286	0.359	-0.074***	0.028
	(0.436)	(0.459)	(0.023)	(0.452)	(0.480)	(0.015)	(0.028)
		Difference-in-difference-in-differences: low versus high predicted likelihood					-0.008
							(0.040)

Note: The two panels present the means (and standard deviations) of postsecondary enrollment (top panel) and persistence beyond second year in college (bottom panel) for students 0-10 points below and above the cutoff broken down by regime; differences between discretion and rule regimes for low and high predicted remediation likelihood groups (low ≤ 0.3 and high > 0.3); difference in these differences between students below and above the cutoff, and difference-in-difference-in-differences between high and low predicted likelihood students. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 9: Discretion versus No Test: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results

	2-year degree attainment						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	No test regime	Difference	Discretion regime	No test regime	Difference	Difference-in-differences
High predicted likelihood	0.172	0.176	-0.004	0.213	0.250	-0.037**	0.033
	(0.377)	(0.381)	(0.021)	(0.409)	(0.433)	(0.017)	(0.027)
Low predicted likelihood	0.228	0.223	0.005	0.222	0.261	-0.039***	0.044*
	(0.420)	(0.416)	(0.022)	(0.415)	(0.439)	(0.014)	(0.026)
			Difference-in-difference-in-differences: low versus high predicted likelihood				-0.012
							(0.037)
	4-year degree attainment						
	0-10 points below the cutoff			0-10 points above the cutoff			
	Discretion regime	No test regime	Difference	Discretion regime	No test regime	Difference	Difference-in-differences
High predicted likelihood	0.0520	0.0541	-0.002	0.102	0.104	-0.003	0.001
	(0.222)	(0.226)	(0.013)	(0.302)	(0.306)	(0.012)	(0.017)
Low predicted likelihood	0.0683	0.116	-0.048***	0.118	0.153	-0.035***	-0.013
	(0.252)	(0.321)	(0.015)	(0.322)	(0.360)	(0.011)	(0.019)
			Difference-in-difference-in-differences: low versus high predicted likelihood				0.014
							(0.026)

Note: The two panels present the means (and standard deviations) of 2-year degree attainment (top panel) and 4-year degree attainment (bottom panel) for students 0-10 points below and above the cutoff broken down by regime; differences between discretion and rule regimes for low and high predicted remediation likelihood groups (low ≤ 0.3 and high > 0.3); difference in these differences between students below and above the cutoff, and difference-in-difference-in-differences between high and low predicted likelihood students. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 10: Discretion versus No Test: Effects on Intermediate Outcomes in High School

	DDD estimates
Following year	
Placed in remedial schedule in ELA	0.390***
	(0.024)
Classroom-peer prior test scores: ELA courses	-0.139***
	(0.043)
Classroom-peer prior test scores: other core courses	-0.035
	(0.041)
Classroom-peer subsidized meal eligible: ELA courses	0.028
	(0.020)
Classroom-peer subsidized meal eligible: other core courses	0.018
	(0.019)
Teacher value-added score in ELA courses	-0.005
	(0.004)
High school outcomes	
Number of advanced courses taken	-0.073
	(0.404)
Number of AP courses taken	0.046
	(0.386)
Ever passed an AP test	0.001
	(0.036)
Number of AP tests passed	0.047
	(0.090)
Obtained a high school diploma	-0.036
	(0.035)
Number of students	10,001

Note: The first column presents the difference-in-difference-in-differences (DDD) estimates, comparing high versus low predicted remediation likelihood groups (low ≤ 0.3 and high > 0.3), below versus above the cutoff, and discretion versus no test regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The second column presents the difference-in-differences in regression discontinuity (DD in RD) estimates, comparing the discontinuity in the outcome of interest at the remediation cutoff for high versus low predicted remediation likelihood students under discretion versus rule regimes. DD in RD coefficients are obtained parametrically, using a linear polynomial and a bandwidth of 10 points. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 11: Discretion versus No Test: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by Student Attributes

		Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
White		-0.001	0.059	0.062	0.000	0.332***
	N=3,386	(0.081)	(0.073)	(0.066)	(0.052)	(0.043)
Black		-0.042	-0.040	0.001	0.037	0.353***
	N=2,365	(0.091)	(0.075)	(0.068)	(0.047)	(0.049)
Hispanic		-0.052	-0.100	-0.112*	-0.004	0.450***
	N=3,444	(0.076)	(0.065)	(0.063)	(0.036)	(0.038)
Always subsidized meal eligible		-0.045	-0.063	-0.057	0.039	0.366***
	N=4,447	(0.066)	(0.055)	(0.052)	(0.031)	(0.034)
Never subsidized meal eligible		0.027	-0.019	0.007	-0.069	0.371***
	N=2,559	(0.095)	(0.091)	(0.084)	(0.065)	(0.052)
English non-native		-0.024	-0.133*	-0.152**	0.004	0.411***
	N=2,772	(0.087)	(0.077)	(0.075)	(0.044)	(0.043)
English native		-0.013	0.040	0.043	0.017	0.383***
	N=7,208	(0.053)	(0.046)	(0.043)	(0.031)	(0.028)
Male		0.014	-0.002	-0.045	0.027	0.392***
	N=4,956	(0.063)	(0.052)	(0.049)	(0.030)	(0.033)
Female		-0.042	-0.015	0.023	-0.000	0.390***
	N=5,024	(0.065)	(0.060)	(0.056)	(0.042)	(0.033)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by student groups, comparing high versus low predicted remediation likelihood groups (low ≤ 0.3 and high > 0.3), below versus above the cutoff, and discretion versus no test regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 12: Discretion versus No Test: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by School Attributes

	Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
% White					
Below median	-0.073	-0.035	-0.075	0.015	0.384***
N=4,040	(0.071)	(0.061)	(0.057)	(0.037)	(0.036)
Above median	0.010	0.032	0.039	0.023	0.379***
N=5,486	(0.062)	(0.054)	(0.050)	(0.036)	(0.031)
% subsidized meal eligible					
Below median	-0.034	-0.010	0.001	0.019	0.325***
N=5,144	(0.064)	(0.057)	(0.053)	(0.038)	(0.033)
Above median	-0.016	0.023	-0.012	0.021	0.422***
N=4,382	(0.068)	(0.058)	(0.054)	(0.035)	(0.034)
ELA scores – mean					
Below median	-0.068	-0.002	-0.003	-0.020	0.392***
N=3,279	(0.076)	(0.062)	(0.058)	(0.035)	(0.042)
Above median	0.006	0.012	-0.001	0.043	0.366***
N=6,247	(0.058)	(0.053)	(0.049)	(0.036)	(0.029)
ELA scores – standard deviation					
Below median	-0.121*	-0.010	-0.028	0.020	0.394***
N=5,154	(0.062)	(0.055)	(0.051)	(0.035)	(0.033)
Above median	0.103	0.035	0.030	0.021	0.337***
N=4,372	(0.070)	(0.061)	(0.057)	(0.040)	(0.034)
% below remediation cutoff					
Below median	0.006	0.039	0.003	0.058	0.321***
N=5,197	(0.065)	(0.058)	(0.054)	(0.039)	(0.032)
Above median	-0.058	-0.029	-0.015	-0.015	0.402***
N=4,329	(0.068)	(0.057)	(0.054)	(0.034)	(0.035)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by school attributes, comparing high versus low predicted remediation likelihood groups (low ≤ 0.3 and high > 0.3), below versus above the cutoff, and discretion versus no test regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, below the cutoff indicator, and the discretion regime indicator. The analysis only includes students that have never been flagged for remediation in the prior two years. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. School characteristics are calculated using student-level data between 2005-06 and 2011-

12 school years. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 13: No Test versus Rule: Effects on Remedial Placement, Postsecondary Enrollment and Persistence, Difference-in-Difference-in-Differences Results

	Placed in ELA remediation						
	Flagged for remediation in prior year			Not flagged for remediation in prior year			
	No test regime	Rule regime	Difference	No test regime	Rule regime	Difference	Difference-in-differences
High predicted likelihood	0.743	0.036	0.707***	0.0335	0.0126	0.021***	0.686***
	(0.437)	(0.186)	(0.016)	(0.180)	(0.112)	(0.007)	(0.017)
Low predicted likelihood	0.655	0.066	0.590***	0.0111	0.0116	-0.001	0.590***
	(0.476)	(0.248)	(0.033)	(0.105)	(0.107)	(0.003)	(0.033)
			Difference-in-difference-in-differences: low versus high predicted likelihood				0.096***
							(0.037)
	Postsecondary enrollment						
	Flagged for remediation in prior year			Not flagged for remediation in prior year			
	No test regime	Rule regime	Difference	No test regime	Rule regime	Difference	Difference-in-differences
High predicted likelihood	0.527	0.494	0.033	0.493	0.568	-0.075***	0.108***
	(0.500)	(0.500)	(0.020)	(0.500)	(0.495)	(0.021)	(0.029)
Low predicted likelihood	0.504	0.502	0.003	0.572	0.590	-0.018	0.021
	(0.501)	(0.500)	(0.039)	(0.495)	(0.492)	(0.012)	(0.041)
			Difference-in-difference-in-differences: low versus high predicted likelihood				0.087*
							(0.050)
	Persisted beyond second year						
	Flagged for remediation in prior year			Not flagged for remediation in prior year			
	No test regime	Rule regime	Difference	No test regime	Rule regime	Difference	Difference-in-differences
High predicted likelihood	0.270	0.240	0.030*	0.288	0.319	-0.031*	0.061**
	(0.444)	(0.427)	(0.018)	(0.453)	(0.466)	(0.019)	(0.026)
Low predicted likelihood	0.248	0.286	-0.038	0.339	0.335	0.005	-0.043
	(0.433)	(0.452)	(0.034)	(0.474)	(0.472)	(0.012)	(0.036)
			Difference-in-difference-in-differences: low versus high predicted likelihood				0.103**
							(0.045)

Note: The two panels present the means (and standard deviations) of postsecondary enrollment (top panel) and persistence beyond second year in college (bottom panel) for students that were flagged for remediation in prior year versus not broken down by regime; differences between no test and rule regimes for high and low predicted remediation likelihood groups (low ≤ 0.35 and high > 0.35); difference in these differences between students flagged and not flagged in the previous school year, and difference-in-difference-in-differences between high and low predicted likelihood students. The analysis only includes students that scored within 10 points above the remediation cutoff. Predicted probabilities are obtained using the "not flagged in prior year" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 14: No Test versus Rule: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results

	2-year degree attainment						
	Flagged for remediation in prior year			Not flagged for remediation in prior year			
	No test regime	Rule regime	Difference	No test regime	Rule regime	Difference	Difference-in-differences
High predicted likelihood	0.220	0.212	0.007	0.237	0.236	0.001	0.007
	(0.414)	(0.409)	(0.017)	(0.425)	(0.425)	(0.018)	(0.024)
Low predicted likelihood	0.178	0.223	-0.045	0.259	0.258	0.001	-0.045
	(0.384)	(0.417)	(0.031)	(0.438)	(0.438)	(0.011)	(0.033)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.052
							(0.041)
	4-year degree attainment						
	Flagged for remediation in prior year			Not flagged for remediation in prior year			
	No test regime	Rule regime	Difference	No test regime	Rule regime	Difference	Difference-in-differences
High predicted likelihood	0.079	0.053	0.026**	0.098	0.121	-0.023*	0.049***
	(0.270)	(0.224)	(0.010)	(0.297)	(0.326)	(0.012)	(0.016)
Low predicted likelihood	0.092	0.084	0.007	0.135	0.116	0.020**	-0.013
	(0.289)	(0.278)	(0.022)	(0.342)	(0.320)	(0.008)	(0.024)
Difference-in-difference-in-differences: low versus high predicted likelihood							0.062**
							(0.029)

Note: The two panels present the means (and standard deviations) of 2-year degree attainment (top panel) and 4-year degree attainment (bottom panel) for students that were flagged for remediation in prior year versus not broken down by regime; differences between no test and rule regimes for high and low predicted remediation likelihood groups (low ≤ 0.35 and high > 0.35); difference in these differences between students flagged and not flagged in the previous school year, and difference-in-difference-in-differences between high and low predicted likelihood students. The analysis only includes students that scored within 10 points above the remediation cutoff. Predicted probabilities are obtained using the "not flagged in prior year" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses for the estimated differences while standard deviations are given under sample means.

Table 15: No Test versus Rule: Effects on Intermediate Outcomes in High School, Difference-in-Difference-in-Differences Results

	DDD estimates
Following year	
Classroom-peer prior test scores: ELA courses	0.133***
	(0.050)
Classroom-peer prior test scores: other core courses	0.132***
	(0.046)
Classroom-peer subsidized meal eligible: ELA courses	-0.061***
	(0.020)
Classroom-peer subsidized meal eligible: other core courses	-0.053***
	(0.020)
Teacher value-added score in ELA courses	-0.003
	(0.003)
High school outcomes	
Number of advanced courses taken	0.590
	(0.464)
Number of AP courses taken	0.027
	(0.437)
Ever passed an AP test	0.010
	(0.037)
Number of AP tests passed	0.057
	(0.094)
Obtained a high school diploma	0.003
	(0.041)
Number of students	16,733

Note: The results present the difference-in-difference-in-differences (DDD) estimates, comparing high versus low predicted remediation likelihood groups (low ≤ 0.35 and high > 0.35), flagged versus not flagged in prior year, no test versus rule regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, flagged in previous year indicator, and the no test regime indicator. The analysis only includes students that scored within 10 points above the remediation cutoff. Predicted probabilities are obtained using the "not flagged in prior year" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 16: Discretion versus No Test: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by Student Attributes

		Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
White		0.031	0.082	0.012	0.103	0.104
	N=5,050	(0.098)	(0.093)	(0.083)	(0.065)	(0.077)
Black		0.158*	0.117	0.102	-0.016	0.166**
	N=4,040	(0.094)	(0.080)	(0.070)	(0.057)	(0.070)
Hispanic		0.080	0.077	0.037	0.057	0.102*
	N=5,796	(0.081)	(0.069)	(0.066)	(0.038)	(0.060)
Always subsidized meal eligible		0.156**	0.106*	0.046	0.027	0.148***
	N=7,206	(0.071)	(0.060)	(0.056)	(0.036)	(0.051)
Never subsidized meal eligible		0.018	0.091	0.055	0.101	0.052
	N=4,298	(0.108)	(0.113)	(0.100)	(0.089)	(0.091)
English non-native		0.116	0.094	0.071	0.019	0.083
	N=4,815	(0.085)	(0.076)	(0.073)	(0.047)	(0.060)
English native		0.065	0.098*	0.038	0.079**	0.105**
	N=11,354	(0.062)	(0.055)	(0.049)	(0.037)	(0.047)
Male		0.073	0.121**	0.121**	0.013	0.120**
	N=7,941	(0.072)	(0.059)	(0.053)	(0.038)	(0.056)
Female		0.123*	0.100	-0.002	0.115***	0.071
	N=8,228	(0.069)	(0.064)	(0.060)	(0.043)	(0.049)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by student groups, comparing high versus low predicted remediation likelihood groups (low ≤ 0.35 and high > 0.35), flagged versus not flagged in prior year, no test versus rule regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, flagged in previous year indicator, and the no test regime indicator. The analysis only includes students that scored within 10 points above the remediation cutoff. Predicted probabilities are obtained using the "not flagged in prior year" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1 percent level respectively. Robust standard errors are given in parentheses.

Table 17: Discretion versus No Test: Effects on Postsecondary Degree Attainment, Difference-in-Difference-in-Differences Results, Heterogeneous Effects by School Attributes

	Postsecondary enrollment	Persisted beyond second year	2-year degree attainment	4-year degree attainment	Placed in remedial schedule
% White					
Below median	0.092	0.063	0.037	0.020	0.201***
N=7,255	(0.075)	(0.065)	(0.061)	(0.038)	(0.055)
Above median	0.119*	0.133**	0.062	0.096**	0.014
N=8,867	(0.070)	(0.064)	(0.057)	(0.044)	(0.049)
% subsidized meal eligible					
Below median	0.088	0.137**	0.077	0.120***	0.073
N=8,277	(0.073)	(0.068)	(0.061)	(0.045)	(0.055)
Above median	0.110	0.060	0.021	0.004	0.133***
N=7,845	(0.071)	(0.061)	(0.057)	(0.039)	(0.048)
ELA scores – mean					
Below median	0.108	0.014	-0.010	-0.006	0.192***
N=5,993	(0.080)	(0.067)	(0.060)	(0.042)	(0.056)
Above median	0.078	0.142**	0.076	0.100**	0.048
N=10,129	(0.066)	(0.061)	(0.057)	(0.041)	(0.048)
ELA scores – standard deviation					
Below median	0.149**	0.118*	0.084	0.008	0.040
N=8,875	(0.068)	(0.061)	(0.056)	(0.041)	(0.046)
Above median	0.021	0.075	-0.007	0.132***	0.151***
N=7,247	(0.077)	(0.070)	(0.064)	(0.044)	(0.057)
% below remediation cutoff					
Below median	0.093	0.167**	0.090	0.126***	0.020
N=8,473	(0.071)	(0.066)	(0.060)	(0.044)	(0.050)
Above median	0.106	0.017	0.001	-0.014	0.185***
N=7,649	(0.074)	(0.064)	(0.057)	(0.040)	(0.055)

Notes: Each number presents the difference-in-difference-in-differences (DDD) estimates broken down by school attributes, comparing high versus low predicted remediation likelihood groups (low ≤ 0.35 and high > 0.35), flagged versus not flagged in prior year, no test versus rule regime. The DDD estimate provided represents the coefficient on the interaction between high remediation probability indicator, flagged in previous year indicator, and the no test regime indicator. The analysis only includes students that scored within 10 points above the remediation cutoff. Predicted probabilities are obtained using the "not flagged in prior year" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors. *, **, *** imply that the estimated difference is statistically different than zero at 10, 5, and 1

percent level respectively. Robust standard errors are given in parentheses. School characteristics are calculated using student-level data between 2005-06 and 2011-12 school years.

Online Appendix Table 1: Average Predicted Remediation Probabilities under the Discretion Regime, Students Never Flagged for Remediation, 0-10 points Below and Above the Cutoff

	Below the cutoff	Above the cutoff
Rule regime	0.515	0.406
	(0.503)	(0.483)
No test regime	0.305	0.243
	(0.324)	(0.326)
Discretion regime	0.274	0.236
	(0.242)	(0.351)

Note: This table presents the average predicted remediation probability under the discretion regime for students never flagged for remediation in the prior two years below and above the cutoff. Predicted probabilities are obtained using the "never-flagged" students 0-10 points below the cutoff under the discretion regime using student test score history (excluding prior year ELA score), course history, and prior disciplinary incidents and suspensions as predictors.