

The Unintended Consequences of Academic Leniency*

A. Brooks Bowden Viviana Rodriguez Zach Weingarten

September 2, 2025

Abstract

Education is widely recognized as the primary driver of human capital accumulation. However, policies designed to promote educational attainment often overlook students' endogenous responses. This paper provides empirical evidence of the causal impact of relaxed academic standards on student effort and achievement. We find that academic leniency results in mechanical GPA increases, but significant reductions in effort among lower ability students based on lower attendance and lower numerical grades. These heterogeneous effects compound and exacerbate gaps in engagement and achievement throughout high school, ultimately worsening deficits in human capital accumulation as measured by ACT scores.

JEL Classifications: I24, J24, I28

Keywords: Academic Standards, Student Effort, Human Capital, Educational Gaps

*We would like to thank Stephen Ross, a second editor, and three anonymous reviewers for their thoughtful and constructive comments. We would also like to thank Samson Alva, Drew Bailey, Shaun Dougherty, John Friedman, Alper Hayri, Brian Jacob, Jonathan Moreno-Medina, Evan Riehl, Petra Todd, participants at the Southern Economic Association, the Advanced Quantitative Methods and Analytics for Public Policy Support, the Causal Inference in Education Research Seminar, the Association for Education Finance and Policy, the UTSA Brown Bag, and the CBCSE Brown Bag for helpful comments and suggestions. Anika Alam and Hannah Lee provided excellent research assistance. This paper was supported by the National Science Foundation Graduate Research Fellowship Program and the Institute of Education Sciences, U.S. Department of Education, through Grant R305B200035 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This project was approved by the University of Pennsylvania Institutional Review Board under protocol #844746. All remaining errors are our own. Contact information: Bowden: University of Pennsylvania, bbowden@upenn.edu. Rodriguez: University of Texas at San Antonio, viviana.rodriguez@utsa.edu. Weingarten: University of Pennsylvania, zachwein@upenn.edu.

I Introduction

Education is often considered “the great equalizer” for its unique capacity to promote social mobility through the accumulation of human capital. In response to widening socioeconomic achievement gaps and increased demand for high quality post-secondary education, state and federal governments have enacted policies to promote educational attainment. These efforts have had reported success. The last decade has seen steady increases in graduation rates and grade point averages in U.S. high schools, yet student achievement—as measured by NAEP, ACT, and SAT scores—has stagnated. This, coupled with decreasing college enrollment rates, suggests a decline in academic standards (Murnane, 2013; Blagg and Chingos, 2016; Hurwitz and Lee, 2018; Harris et al., 2023). At face value, a decline in academic standards for graduation can possibly lead to more equitable outcomes because a larger share of students meet graduation requirements, obtain a high school diploma, and gain access to post-secondary education opportunities.¹ However, this simplified view ignores how students may endogenously respond to changing academic standards, which can directly impact their human capital accumulation.²

This paper explores the relationship between academic standards, student effort, and human capital accumulation by analyzing a statewide policy that abruptly lowered grading standards for high school students. We recover the causal effect of this change on student outcomes by constructing a mechanism that leverages the implementation timing of the policy in conjunction with a separate assignment rule that determines high school entry eligibility. In doing so, we are the first to recover causal effects of lowered standards on students’ academic and behavioral outcomes in a U.S. setting.

Understanding the effects of academic standards on student outcomes is challenging because standards are intrinsic to teachers, schools, and districts. As students select teachers and schools, the effects of grade leniency on student outcomes become confounded by sorting (Figlio and Lucas, 2004). Furthermore, changes in academic standards often occur gradually over time, making it impossible to disentangle treatment effects from unobserved time trends. We overcome these challenges by exploiting two sources of variation.

First, we exploit a North Carolina policy that lowered high school grading standards. In the

¹For example, Landaud et al. (2024) find that students who obtain an exogenous boost in GPA can access more selective higher education programs that have higher returns. These effects later translate into better adult labor market earnings.

²In a similar vein, this viewpoint also disregards the downstream effects that increased leniency may have on post-secondary and labor markets that respond to the now weaker signal-to-noise ratio associated with a high school diploma.

Fall of 2014, the North Carolina State Board of Education changed high school grading policy from a 7-point scale to a 10-point scale to standardize grading practices and increase the competitiveness of students applying to colleges. This policy effectively decreased the associated numeric threshold for all letter grades for *all* students across all subjects, courses, and teachers, thereby moving all public high schools to a more lenient grading scale. This institutional setting provides a unique opportunity to observe a stark, ubiquitous change in academic leniency.

Our second mechanism (quasi-)randomly assigns students to the stricter grading standard or the more lenient one. In North Carolina, schools follow a birth date cut off to assign students into kindergarten, which later extends to high school entry assignment as students matriculate in 9th grade. We combine both sources of variation using a difference-in-discontinuity design to identify the causal effect of academic leniency.

We rely on rich administrative data on the universe of North Carolina public school students from 2013 to 2019, with information on demographics, school engagement, and learning during high school. Using exact birth date records from the North Carolina Department of Public Instruction, we generate a measure of relative distance to the annual kindergarten entry rule cut-off. This distance measure is the running variable in our difference-in-discontinuity specification. We find that increased leniency resulted in large effort reductions as measured by increases in absences (22%) and reductions in numerical grades in core subjects of English and math (2%), with no changes in the distribution of standardized test scores. As expected, we find that GPA increases (4.8%), but we show that this is largely mechanical.

As prior performance may relate to how students respond to increased leniency, we explore heterogeneity across observed student ability, as measured by 8th grade Math performance. The results show stark differences in student response. Students at the top of the ability distribution experienced no change in attendance, while students at the middle and bottom of the distribution drove the rise in absenteeism with large increases in the number of school days missed. Despite these differences, we find that GPA gains are very similar across ability groups. This result is striking given that the policy’s design mechanically increased GPA for all students but especially for students at the margin of passing or failing a course. Therefore, *a priori*, we would expect GPA gains to be concentrated among students on the lower end of the ability distribution. Our opposing finding, together with heterogeneous increases in absences, suggests a negative effort response among middle and low ability groups that undoes mechanical gains brought on by the policy change. However, we are unable to rule out potential confounding effects on GPA that may

be related to teacher response.³

A negative effort response, especially through increased absences, could increase the likelihood that students would exit high school with low levels of human capital. Thus, we explore outcomes among our 9th grade cohort throughout high school to better understand the effects of grading leniency over time. We find widening gaps in long-run effort with absences continuing to increase in later grades, particularly among low and medium ability students who miss an additional 3-5 days of school in 11th grade. Furthermore, we find that the mechanical gains in 9th grade GPA are eliminated by 10th grade, which yield net losses to GPA in 11th grade. However, despite starkly heterogeneous and compounding absence effects across ability groups, we are unable to detect differences across ability groups in long-term GPA effects.

To understand how long-run effects on absences may translate into longer-run human capital accumulation, we draw from student performance on the ACT college entrance exam. An attractive feature of using these data is that all 11th graders are required to take the ACT, allowing us to recover a nationally standardized measure of cognitive ability that does not suffer from selection bias. We find that ACT scores decline by 0.5 points (2.4%) on average. Consistent with other findings, we find suggestive evidence indicating a reduction in ACT scores driven by students at the middle and bottom of the ability distribution. These longer-run results suggest that academic leniency exacerbated achievement gaps and lowered human capital accumulation for students who may already be at a deficit.

Finally, we explore whether our results for 9th graders generalize to older cohorts of students. We find that the first-year effects for 10th, 11th, and 12th graders are consistent, albeit somewhat attenuated, with our main findings. Across all cohorts, we find that the policy increased both GPA and absences, with the latter being smaller and more noisily estimated than our main effects. In contrast to the results for our 9th grade cohort, we do not find that the policy had an immediate or long-term effect on college-readiness proxies for older cohorts of students. These null effects might be indicative of both the pivotal role that 9th grade can have on establishing norms in high school and the differences in exposure to lenient grading across cohorts.

Our results are robust to a battery of checks including placebo exercises, tests for manipulation of the running variable and covariate smoothness, changes to the selection of control cohort windows, bandwidth selection, spurious changes to relevant educational inputs, and model specification. We

³ For example, teachers might be less likely to round at the pass threshold after the reform, which would suppress numeric grades and erode GPA effects for students at the bottom of the distribution.

do not find any evidence that these potential sources of influence meaningfully impact our estimated results.

This paper contributes to a relatively large literature on grading standards and its effects on student welfare, particularly academic achievement.⁴ Two papers are especially relevant to our study. First, [Hvidman and Sievertsen \(2021\)](#) leverage variation from a 2007 Danish grading reform that recoded high school students' grade point averages. The authors show that students who randomly received lower grades as a result of this change in grading scheme obtained higher levels of achievement in subsequent years — as measured by both higher GPA and test scores — and enrolled in university at higher rates. Given that this recoding did not convey information about students' actual performance, these gains in achievement are likely driven by students' effort response to the policy, but the authors cannot directly test for this. In their setting, [Hvidman and Sievertsen \(2021\)](#) find that higher achieving students perform better than their peers when they receive a negative shock. Suggesting that high-performing students respond to this policy with increased levels of effort.

Second, [Butcher, McEwan and Weerapana \(2024\)](#) examine the 2014 implementation of pass/fail grading for first-year, first-semester students at Wellesley College. The authors find that the policy led students to obtain lower numeric grades, equivalent to 23% of a standard deviation reduction in the *implicit* GPA, as pass/fail grades are not accounted for in GPA calculations. The authors rule out several potential mechanisms driving this effect and conclude that the likely mechanism is decreased levels of student effort, although they cannot directly test for it. The reduction in mean grades is greater among students with below-median math skills that do not have any financial aid.⁵

Our findings are broadly consistent with [Butcher, McEwan and Weerapana \(2024\)](#), showing that relaxing grading standards leads to an overall reduction in student effort particularly for students on the lower end of the ability distribution. While our paper does not directly test the inverse setting, [Hvidman and Sievertsen \(2021\)](#) find that more stringent academic standard elicits higher effort levels from students, with effects concentrated among high ability students.

⁴ Other work has examined the impact of grading standards on how students sort into courses and programs ([Bar, Kadiyali and Zussman, 2009](#); [Butcher, McEwan and Weerapana, 2014](#); [Ahn et al., 2024](#)), college graduation rates ([Denning et al., 2022](#)), and how students perform in the labor market ([Hansen, Hvidman and Sievertsen, 2023](#); [Landaud et al., 2024](#)).

⁵ Another relevant study is [Figlio and Lucas \(2004\)](#) that examines grading standards at the teacher level by leveraging a large panel of teacher-student matched data with information on test score gains and its relationship with numeric grades assigned by teachers. The authors find that students who take classes with teachers who have stringent grading standards have higher levels of subsequent achievement. The authors also find important heterogeneity across student ability with high-ability students experiencing the largest benefit from high standards.

Our most important contribution to this literature is to directly uncover the mechanism that ties grading standards to subsequent student academic achievement in these settings: student effort, as measured by attendance. Student effort is a key contributor to academic success (Durden and Ellis, 1995; Stinebrickner and Stinebrickner, 2008; Metcalfe, Burgess and Proud, 2019). Our findings that lenient academic standards lead to a reduction in effort, ultimately generating decreases in achievement as measured by ACT scores, further supports this body of work.⁶

The remainder of the paper is structured as follows: Section II describes the grading policy change implemented by the North Carolina State Board of Education in 2014. Section III describes the data used in the analysis and provides summary statistics for our final analysis sample. Section IV establishes the research design. Section V presents our results, which are validated in Section VI. Finally, Section VII concludes the paper.

II Institutional Background

II.A Standardization of High School Grading Policies

Table 1: Changes in Academic Course Grades

Letter Grade (Grade Point)	Original	New	Δ Min. Requirement
A (4.0)	93 – 100	90 – 100	–3
B (3.0)	85 – 92	80 – 89	–5
C (2.0)	77 – 84	70 – 79	–7
D (1.0)	70 – 76	60 – 69	–10
F (0.0)	0 – 69	0 – 59	

NOTES: The above table displays the changes in grade thresholds as a result of the policy change. The first column displays the associated letter grade and points (used to calculate GPA) for a given threshold. “Original” refers to the standards mandated by the state prior to the 2015-2016 school year, and “New” refers to the updated ones. Δ Min. Requirement displays the point difference generated by the policy for each letter grade.

We focus our empirical analysis on high school students in North Carolina. Unlike most states in the U.S., the North Carolina State Board of Education explicitly outlines grading standards for all public high schools in the state. These standards include grading scales that reflect the correspondence of numeric scores to letter grades. In the Fall of 2014, the North Carolina State Board of Education voted to standardize high school grading policies to a 10-point scale in an effort

⁶ Our work also relates to the literature on the unintended consequences of tightening accountability measures for public schools (Koretz, 2002; Jacob, 2005; Glewwe, Ilias and Kremer, 2010; Jacob, 2005; Figlio, 2006; Cullen and Reback, 2006; Dee et al., 2016). Our paper speaks to the unintended response of policies in shaping student behavior. Consistent with this literature, we find a concentration of gains for students at the top of the ability distribution with negative impacts for students at the bottom end of the distribution.

to increase comparability between school districts and increase the competitive quality of students applying to colleges.⁷ Importantly, this policy applied to all public high school students in the state, across all high school grades.

Table 1 outlines the specific changes associated with each letter grade. The change in letter grade standards created an additional 10-point buffer at the margin of passing a class. For example, a student taking a math class in the 2014-2015 school year would need a 70 or higher to earn credit for the course, while a student in that same class the following year would instead need a minimum grade of 60. The final column in Table 1 displays the relative reduction in standards associated with each letter grade. As shown, leniency increased most for students at the lower tail of the grade distribution.

Students may also experience changes in grading standards through teacher grading practices. To assess the extent to which state-mandated changes in grading standards led to changes in grading practices, Figure 1 displays histograms of 9th grade math course grades for the first year of the policy (2016) and the year prior (2015). For all letter grade thresholds, these distributions show bunching just to the right of letter grade cutoffs, especially at the margin of passing a course. The comparison of 2015 to 2016 histograms shows how the distribution of numerical grades immediately shifted in response to the 2016 change in grading standards. We take this as evidence that letter grade assignment is endogenous to statewide standards.

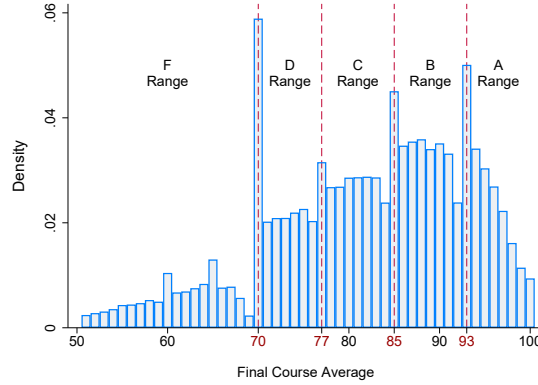
We highlight an important but subtle point about the interpretation of Figure 1. Although the visual evidence clearly shows an endogenous response to statewide standards, we cannot attribute this response to either teachers or students in isolation. In particular, the change in bunching patterns is driven in part by teachers' willingness to round marginal students, which may have differentially changed in response to the 10-point scale. Teachers may be more willing to round a 59 to a 60 than they counterfactually were to round a 69 to a 70. Indeed, Figure 2 shows that the rate of failing grades in core academic courses declined between by 20% with the introduction of the new grading policy.⁸

As a result, any estimates of the policy's effect on student GPA will not have a straightforward interpretation. To see this, first note that the policy mechanically increased GPAs for all students.

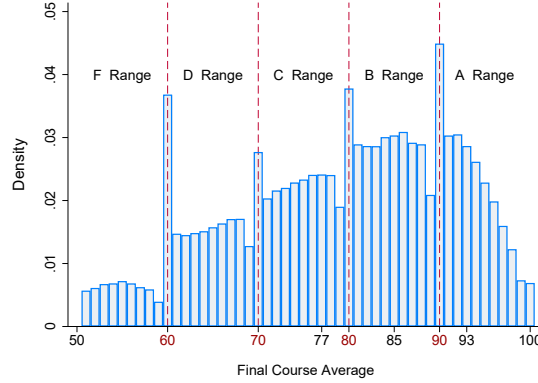
⁷Two months prior, the state additionally approved the adjustment of quality point premiums associated with advanced courses (such as Honors, AP/IB, and college courses) to reflect a maximum GPA of 5.0. In this paper, we will focus on 9th grade student outcomes, therefore, the relevant policy change is the change in letter grade standards. See North Carolina's General Statute 116-11(10a) for more details.

⁸Throughout this paper, we define core academic courses as courses in the subjects of math, English, science, and social studies.

Figure 1: Distribution of 9th Grade Math Final Course Grades



(a) 2015



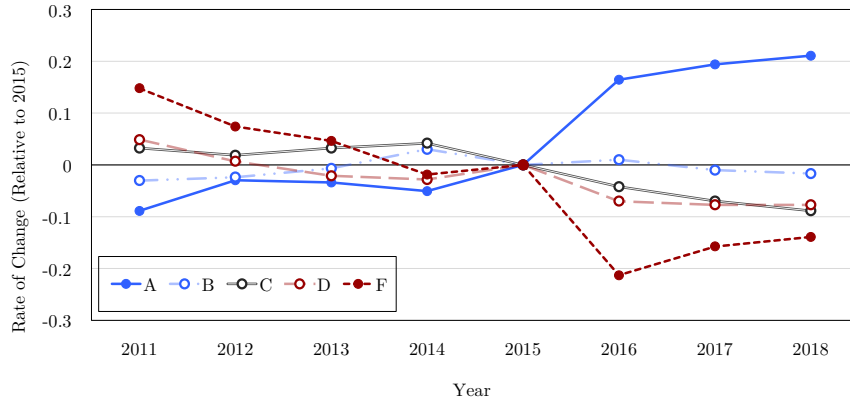
(b) 2016

NOTES: These histograms plot transcript-level math grades for 9th graders in 2015 (panel (a)) and 2016 (panel (b)). Course averages written in red denote the cutoff minimum for each corresponding letter grade; e.g., in the 2014-2015 school year, a 93 corresponds to the lowest numeric course average which earns a grade of A. We censor grades above 100 to have the value 100 and omit grades below 50 from the figures. The distribution for other course types follows a similar pattern. See [Figure B1](#) for equivalent histograms of an additional pre and post treatment year (2014 and 2017).

Second, students may respond to the newfound leniency in grades by endogenously changing their effort. Third, effects on student GPA are likely to be affected by how teachers choose to distribute grades in response to the mandated grading scale change. In an effort to mitigate the influence of this teacher effect, we consider additional outcomes in [Section V](#) that are likely orthogonal to teacher response, namely standardized test scores and absences.

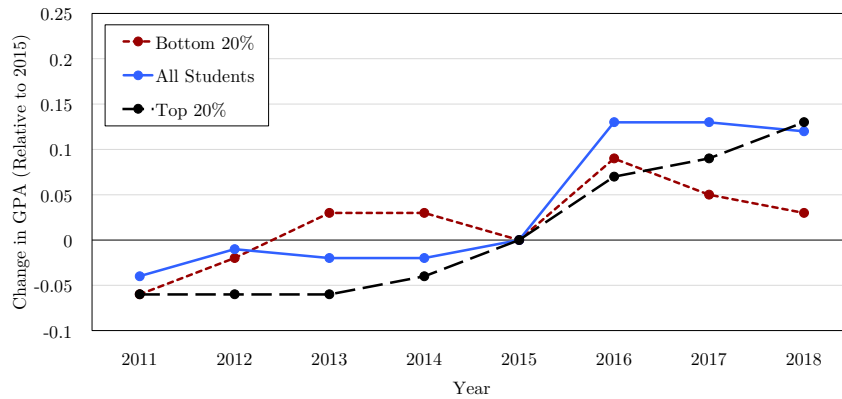
Given the stark shift at the top and bottom of the grade distribution shown in [Figure 2](#), we explore changes in student GPA based on students' prior performance. We use student 8th grade math end-of-grade (EOG) exam scores to proxy for incoming academic preparedness. [Figure 3](#) provides the change in GPA for 9th graders relative to 2015 across student ability. We report average GPA changes for all 9th grade students (in blue), 9th graders belonging to the lowest two

Figure 2: Changes in 9th Grade Letter Grade Shares



NOTES: This figure plots the changes in the proportion of each letter grade earned in core academic courses. We standardize levels relative to 2015, the last year before the policy change occurred. Analogous figures for 10th, 11th, and 12th grade are presented in [Figure B2](#).

Figure 3: Changes in 9th Grade GPA



NOTES: This figure reports relative unweighted 9th grade GPAs in core classes only for students who took the 8th grade math EOG exam, the most-taken EOG for 8th graders. Approximately 90% of all 9th graders have valid scores. The subsamples of students listed above respectively refer to those scoring at or below the 20th percentile in the distribution of EOG scores, all students taking the EOG, and those scoring above the 80th percentile. We standardize levels relative to 2015, the last year before the policy change occurred. Analogous figures for 10th, 11th, and 12th grade are presented in [Figure B3](#).

deciles of 8th grade math EOG score distribution (in red), and 9th graders belonging to the highest two deciles (in black). Despite all students exhibiting immediate gains in GPA, lower achieving students revert back toward pre-policy achievement levels while higher achieving students continue to increase their GPA over time. Taken together, [Figure 3](#) suggests that the primary driver of *sustained* increases to GPA came from the newfound relative ease in earning an A, which translated to GPA gains that were accrued mostly by higher achieving students.

II.B Minimum Age Requirement for School Entry

The second source of identifying variation used in our research design relies on North Carolina’s Minimum Age Requirement for school entry. Under North Carolina’s General Statute 115C-364, children aged 5 years old before October 17th of that school year are entitled to entry into the public school system. Those turning 5 on or after October 17th are required to enroll in the following year. In practice, however, the timing of school entry can be influenced by parents or caregivers. As a result, not all students are perfectly assigned to cohorts based on North Carolina’s minimum age requirement rule. Our empirical specification (see [Section IV](#)) flexibly accounts for noncompliance in kindergarten entry to recover a local average treatment effect (LATE) estimate of the effect of academic leniency on student outcomes.

II.C Importance of 9th Grade

We focus on the first year of high school, 9th grade, to study behavioral responses to academic leniency. As the first year of high school in the U.S., this is an important year for students to form habits and expectations, and to develop knowledge and skills that are critical for future success ([Allensworth and Easton, 2007](#); [Mac Iver et al., 2019](#)). 9th graders also have more autonomy over their coursework and performance compared to earlier years, and their grades and test scores have high stakes implications for graduation, postsecondary, and labor market outcomes. Practically in our setting, 9th grade is the the youngest grade for which all students across the state were expected to experience the grading scale change. Consequently, the 2016 9th grade cohort is the first that we can observe complete exposure to treatment as they matriculate through high school. This year of high school is also convenient because most students take Math I, which is required for graduation and concludes with an end of course (EOC) standardized test, allowing us to estimate the policy’s effect on short-run human capital accumulation. We supplement these analyses with older cohorts in [Section V.D](#).

III Data

We rely on rich administrative data on the universe of North Carolina public school students from the 2012-2013 to 2018-2019 school years provided by the North Carolina Education Research Data

Center (NCERDC).⁹ These data allow us to capture key information on student learning and effort. To examine the effects of grade leniency on student learning, we draw from student performance on Math I standardized tests. Although prior work has used GPA and absences as proxies for student engagement (Hastings, Neilson and Zimmerman, 2012; Jackson, 2018), our policy mechanically shifts GPA in a way that may obscure student response. Thus, we analyze student GPA and absences separately, and we rely primarily on student absences and chronic absenteeism as inverse proxies for student effort.¹⁰ We calculate annual GPA using administrative records of high school transcripts from core courses. Transcript data include the names and statewide codes of courses that students take, corresponding course subject flags, and final numeric mark obtained.

One notable limitation to the data provided by NCERDC is that student birth dates are anonymized at the month level (see Cook and Kang (2016) for a greater discussion). Given that our research design relies on the quasi-random assignment of students to cohorts based on their date of birth, we supplement NCERDC data with restricted-access data on students' exact birth dates provided by the North Carolina Department of Public Instruction.

We make three restrictions to our sample for the purposes of precise empirical estimation. First, we omit students belonging to charter high schools. Although charter schools were mandated to comply with the statewide policy change, they generally have different courses, curricula, and can grade students differently than in the traditional public school system. Their omission therefore ensures comparability of GPAs across students. Second, we drop students with disabilities from the analysis. We do this primarily because these students may not participate in standard 9th grade courses required for graduation, such as Math I. Finally, we restrict our sample to 9th graders for whom we can observe 8th grade Math EOG performance. We draw from 8th grade math scores to classify students into pre-treatment achievement groups and explore heterogeneity across ability. In the interest of comparability across estimates, we restrict our analytic sample to include only students with an 8th grade math test score.

Table 2 presents an overview of the data used in our analysis. Columns (1) and (2) present descriptive statistics for the universe of North Carolina students, while columns (3) and (4) present analogous statistics for the students in our final analytic sample. Students in column (3) make up our treatment group. These students are born within a 180-day bandwidth before or after October

⁹ Throughout this paper we use the spring term year to refer to each school year. We also use publicly available data from North Carolina's Educational Directory and Demographical Information Exchange (EDDIE) to measure school district rurality.

¹⁰ We follow the state's definition of chronic absenteeism as being absent 10% or more of the days in the school, which translates to 18 days.

Table 2: Descriptive Statistics

	Unrestricted Sample		Analytic Sample	
	Treated Birth Cohort (1)	Control Birth Cohorts (2)	Treated Birth Cohort (3)	Control Birth Cohorts (4)
Female	0.503 (0.500)	0.508 (0.500)	0.505 (0.500)	0.511 (0.500)
White	0.508 (0.500)	0.522 (0.500)	0.511 (0.500)	0.526 (0.499)
Asian	0.030 (0.171)	0.028 (0.166)	0.029 (0.168)	0.027 (0.163)
Black	0.253 (0.434)	0.257 (0.437)	0.254 (0.435)	0.258 (0.438)
Hispanic	0.156 (0.363)	0.135 (0.342)	0.152 (0.359)	0.131 (0.338)
Other	0.054 (0.225)	0.057 (0.232)	0.054 (0.226)	0.058 (0.233)
EDS	0.630 (0.483)	0.627 (0.484)	0.642 (0.479)	0.637 (0.481)
Rural	0.479 (0.500)	0.481 (0.500)	0.488 (0.500)	0.491 (0.500)
Observations	102,521	200,886	92,359	181,336

NOTES: This table shows descriptive statistics of student demographic and socioeconomic characteristics. Columns (1) and (2) present descriptive statistics for 9th graders between the 2013-2014 and 2015-2016 academic years, regardless of whether or not they have a valid 8th grade math score. Columns (3) and (4) present descriptive statistics for the subset of 9th graders that have a valid 8th grade standardized math test score on record. Columns (1) and (3) restrict to the birth cohort in the treatment window, i.e., those born no more than 180 days from the kindergarten entry date cutoff for the year that quasi-randomly induces exposure to either the lenient or stringent grading policy. Columns (2) and (4) consider the pooled control birth cohorts for three birth windows prior to the treatment window and three birth windows after.

17th of 2000, the relevant date that sorted students into either the last year of the 7-point scale or the first year of the 10-point scale. Conversely, students in column (4) make up our control group. These students have birth dates in the same window as our treatment group, but in years that are irrelevant for treatment assignment (1998 and 1999). We pool these birth windows and consider them as one control group.

Finally, columns (1) and (2) present analogous statistics to columns (3) and (4) using the unrestricted sample of students. Overall, statistics shown in [Table 2](#) demonstrate a balance of student characteristics between treatment and comparison groups of our analytic sample. When comparing students in our analytic sample to those of the whole state (columns (1) and (2)), we also see no meaningful differences in observable characteristics. With this balance in mind, we use the restricted group as our preferred analytic sample.

IV Research Design

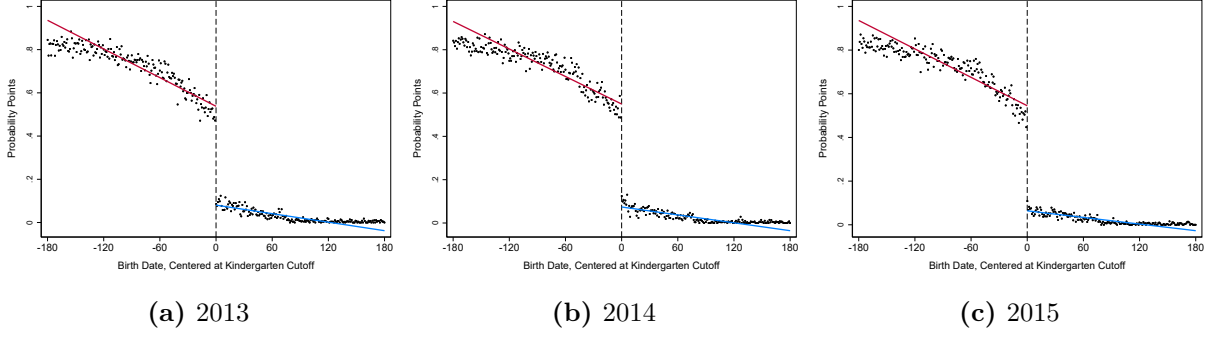
Identifying the causal effect of academic leniency is challenging for several reasons. First, the relaxation of stringent standards often occurs gradually over time. Year-to-year changes in how grades are allocated tend to be small, if not completely indiscernible. This means that testing for sizable effects related to grade inflation requires a long enough panel to capture observable differences in the distribution of grades and student achievement. A simple comparison of cohorts at either end of a sufficiently long panel would exhibit bias induced by unobserved time trends. To circumvent this issue, we exploit the roll-out of a new grading policy in North Carolina that explicitly outlined how schools would distribute letter grades. Importantly, North Carolina issued this change abruptly from one year to the next.

A second issue to consider when analyzing the effects of academic standards is the fact that schools might differ in their explicit grading schemes (e.g., the relative course average needed to earn a letter grade of A) or their leniency *conditional* on the scheme (e.g., the share of A's given). The universality of the policy addresses this first concern directly. Each high school was mandated to follow the 10-point scale regardless of the scale they utilized during the previous academic year. Importantly, we find no evidence to suggest that any school across the state had adopted a 10-point scale prior to this mandate, which provides us with a sharp roll-out and adoption of the new policy.

The third and final challenge, which the policy alone cannot address, is selection into academic cohorts. [Deming and Dynarski \(2008\)](#) show that parents' incentives to "red-shirt" their children has risen over time. This process delays a child's entry into kindergarten with the intended goal of providing them with age advantages during their time in the public school system. The selection on gains induced by this decision has the potential to bias our estimates, especially if parents' strategic incentives differed substantially between the two cohorts of interest. For this reason, a direct comparison of outcomes between the treated (the first cohort to experience the 10-point scale) and the control (the last cohort to experience the 7-point scale) will likely be distorted by endogenous selection bias.

Economists have long exploited exogenous birth date cut-off rules in a regression discontinuity (RD) framework to combat this endogenous enrollment problem ([Angrist and Krueger, 1991](#); [Elder, 2010](#); [Cook and Kang, 2016](#); [Navarro-Palau, 2017](#); [Dee and Sievertsen, 2018](#); [Ordine, Rose and Sposato, 2018](#); [Persson, Qiu and Rossin-Slater, 2021](#)). However, this empirical strategy inherently compares the youngest students in a cohort (those born just to the left of the cut-off) to the oldest

Figure 4: Discontinuity in High School Entry



NOTES: The above figure plots the discontinuity in the probability of entering high school in the stated year for students born no more than 180 before or after the kindergarten cut-off date, October 17th, normalized to zero.

students in another one (those born just to the right). As a result, a standard birthday RD cannot disentangle *age effects* from the desired *policy effects*.

With this in mind, we build on a body of work that combines the traditional birthday RD with a simple difference-in-differences model to construct a difference-in-discontinuity (diff-in-disc) estimator (Grembi, Nannicini and Troiano, 2016; Bertrand, Mogstad and Mountjoy, 2021; Asker, 2024; Behrman et al., 2024). This empirical design leverages the randomization of the kindergarten entry cut-off (October 17th) while separately isolating RD-induced age effects from any direct effects of the policy. Formally, we estimate the following specification,

$$y_i = \beta_0 + \beta_1 p_i + \beta_2 d_i + \tau(p_i \times d_i) + b_i \left(\beta_3 + \beta_4 p_i + \beta_5 d_i + \beta_6 (p_i \times d_i) \right) + b_i^2 \left(\beta_7 + \beta_8 p_i + \beta_9 d_i + \beta_{10} (p_i \times d_i) \right) + \gamma' \mathbf{x}_i + \varepsilon_i, \quad (1)$$

where y_i is some outcome of interest for 9th grade student i , p_i is the “post” indicator, which takes the value 1 if i belongs to the policy-relevant birth cohort (i.e., they are born ± 180 days from the birth date cut-off in 2000, meaning they are quasi-randomly assigned to begin 9th grade in either the last year of the 7-point scale or the first year of the 10-point scale), d_i is an enrollment indicator equal to 1 if i enrolls in 9th grade in year $t + 1$ for a given birth cohort and 0 if they enroll in year t , the running variable b_i is a student’s birthday measured as the distance from the relevant cut-off (normalized to 0), $\mathbf{x}_i$ is a vector of student-level covariates included to improve precision, and ε_i is i.i.d. idiosyncratic error.¹¹

We allow each parameter to vary at the margin of birth date, as first introduced by Grembi, Nan-

¹¹ We include in each specification controls for race, gender, and socioeconomic status.

nicini and Troiano (2016). As written, Equation 1 implies a sharp discontinuity where enrollment in 9th grade is as-good-as-random for any given birth cohort. However, Figure 4 shows a discontinuous jump at the cut point of approximately 45 percentage points each year, suggesting large amounts of persistent noncompliance. Therefore, we implement a fuzzy difference-in-discontinuity design in the spirit of Behrman et al. (2024) by instrumenting for d_i using $b_i^+ \equiv \mathbb{1}_{\{b_i \geq 0\}}$, an indicator for being born on or after October 17th in the corresponding year.¹² By doing so, we control for noncompliance induced by either delayed entry into kindergarten or retention between kindergarten and 9th grade. Additionally, the visual evidence in Figure 4 implies a quadratic fit in the running variable.¹³ To account for this, our main specification includes an additional interaction for each term with b_i^2 . Our parameter of interest is τ , the fuzzy difference-in-discontinuity estimator, which carries a local average treatment effect interpretation.¹⁴ By controlling for age effects through the instrumented d_i , as well as by controlling for spurious cohort effects through p_i , we can more confidently take estimates of τ as capturing the causal effect of the lenient grading reform in isolation.

We estimate Equation 1 using a three-year pooled sample of 9th graders: two control birth cohorts and one “treatment” birth cohort. The control cohorts are born ± 180 days from October 17th in either 1998 or 1999, whereas the treatment birth cohort is born around October 17, 2000.¹⁵ The selection of 180 days as the bandwidth guarantees that no student belongs to multiple birth cohorts while still retaining as much information from the data as possible. We further show in Section VI that our results do not meaningfully depend on either the bandwidth or control group selections.

¹² We instrument for each instance of d_i using the analog term with b_i^+ as a substitute. For example, the term $p_i \times d_i$ is instrumented for by $p_i \times b_i^+$.

¹³ Our results are robust to a linear specification. See Panel D of Table A2.

¹⁴ Table A1 compares attributes of compliers to those of the main sample using methodology in Słoczyński, Uysal and Wooldridge (2024). Although we maintain a local average treatment effect interpretation, we find that the compliers are observably similar to the average student.

¹⁵ We depart from Bertrand, Mogstad and Mountjoy (2021) in including post-policy birth cohorts as the control group in our preferred specification because the policy allowed middle schools – but did not require them – to adopt a 10-point grading scale. As a result, middle school take-up was staggered and non-universal. Our data do not allow us to observe which middle schools adopted the policy, nor when. Thus, pooling over the post-period in this manner would include students in the control that have an equal amount of exposure to the new policy as the treated cohort. Even so, we include a version of the model with post-policy years in Section VI and find only a small level of attenuation in our main results.

V Results

V.A Visual Evidence of Treatment

Figure 5 provides visual evidence to motivate our research design. We plot regression discontinuity bin-scatter plots for student outcomes in control ($p_i = 0$) and treatment ($p_i = 1$) cohort windows. Observed discontinuities at the threshold of control windows can be interpreted as potential age effects that would bias estimates from a simple regression discontinuity design. The discontinuity in the treatment window is therefore a combination of this age effect and the underlying policy effect. Therefore, the difference in discontinuous jumps between control windows and our treatment window can be interpreted as the effect of the grading policy net of any age effects.

Overall, these plots indicate that the policy generated increases in student GPA, decreases in course grades, and increases in absences.¹⁶ Panel A of Figure 5 presents plots for student GPA. These plots show that while there is a positive discontinuity at the threshold for control window cohorts (i.e. older-for-grade students tend to have higher GPAs), the discontinuity at the threshold for our treatment window is much larger. Panels B and C display analogous plots for end-of-semester course grades in 9th grade math and English. In control years, old-for-grade students earn higher grades in math and English, but during the policy shift these gains are completely eliminated. Panel D shows analogous plots for student absences. For this outcome, control window cohorts show a negative discontinuity at the threshold (i.e. older-for-grade students tend to have lower absences). However, the treatment window plot shows a *positive* discontinuity, suggesting large increases in absences driven by the policy.

V.B Causal Effects on 9th Grade Student Outcomes

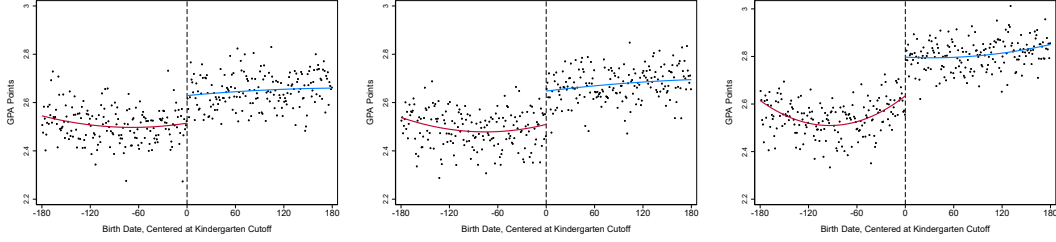
In this section, we present estimated effects of the policy on 9th grade student outcomes. We first examine the effects of grading leniency on student GPA and 9th grade final course averages in math and English, which are direct outcomes of the policy change. However, interpreting effects on grades is challenging because they can partially reflect both mechanical increases driven by the policy and latent responses to the policy from students and teachers. In an attempt to identify student response in isolation, we further explore effects on student learning and engagement through Math I EOC scores and absences in 9th grade, neither of which are manipulable by teachers.

Column (1) of Table 3 reports fuzzy difference-in-discontinuity estimates of Equation 1 for all

¹⁶ See Figure B4 for bin-scatter plots for the likelihood of chronic absence.

Figure 5: Regression Discontinuity of Outcomes Across Academic Years

Panel A. 9th Grade GPA

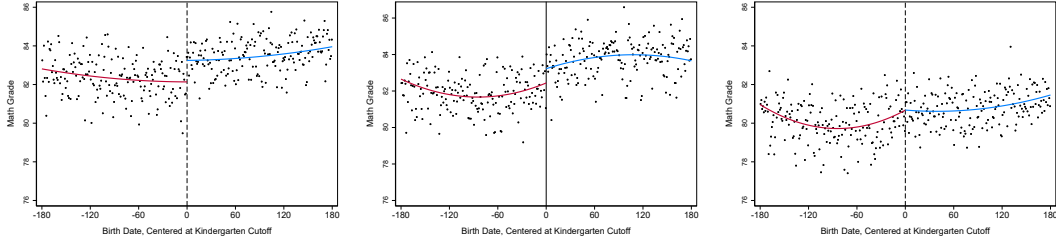


(a) 2014 (Control $p_i = 0$)

(b) 2015 (Control $p_i = 0$)

(c) 2016 (Treatment $p_i = 1$)

Panel B. Math Course Grades

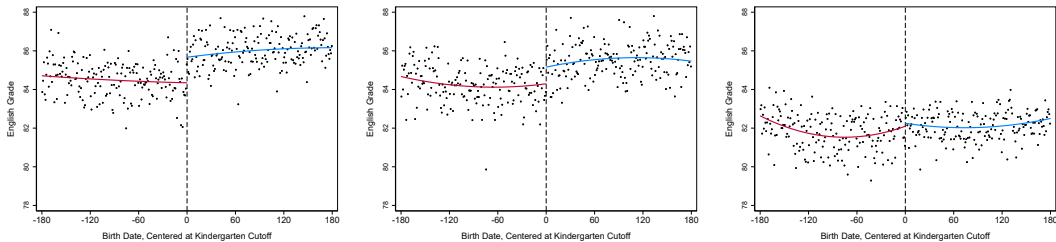


(d) 2014 (Control)

(e) 2015 (Control)

(f) 2016 (Treatment)

Panel C. English Course Grades

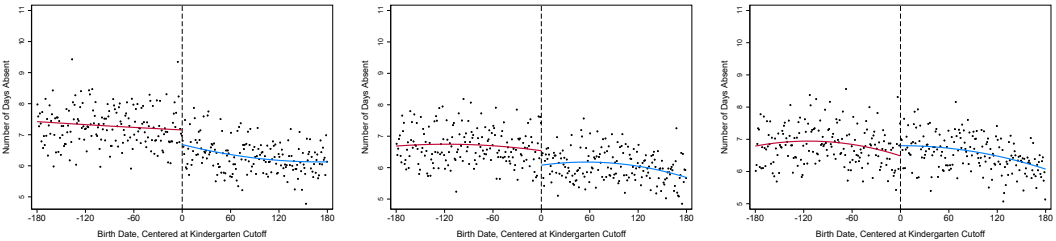


(g) 2014 (Control)

(h) 2015 (Control)

(i) 2016 (Treatment)

Panel D. 9th Grade Absences



(j) 2013-14 (Control)

(k) 2014-15 (Control)

(l) 2016 (Treatment)

NOTES: The above figure shows discontinuities in 9th grade GPA, Math I numeric marks, Math I EOC scores and 9th grade absences (panels A through D) in each listed academic year using a 180-day bandwidth and a quadratic fit. Analogous plots for 9th grade math grades and the likelihood of 9th grade chronic absence is presented in [Figure B4](#).

students and for all outcomes. These estimates indicate that greater grade leniency resulted in an increase of 0.127 GPA points, corresponding to a 4.8% increase relative to the mean. However, higher levels of GPA were not accompanied by greater levels of student achievement, as measured

Table 3: Fuzzy Difference-in-Discontinuity Estimates for 9th Grade Outcomes

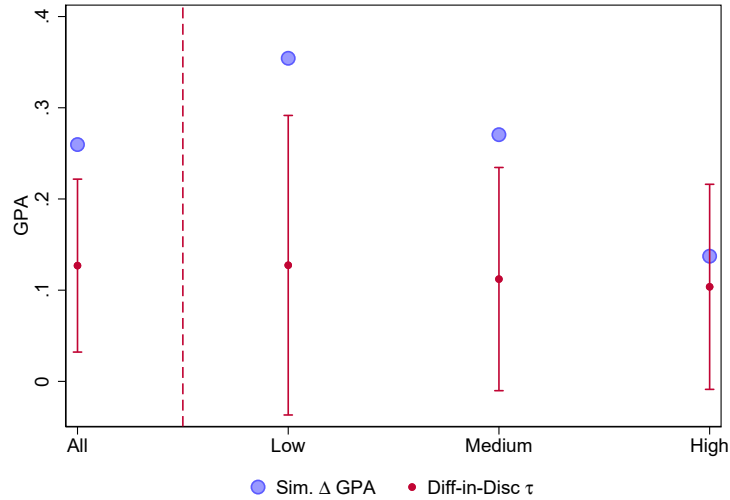
	Analytic Sample (1)	<i>Ability Level</i>		
		Low (2)	Medium (3)	High (4)
Core Academic GPA	0.127 (0.048)	0.127 (0.084)	0.112 (0.062)	0.104 (0.057)
Outcome Mean	2.640	1.722	2.490	3.244
Math Course Grade	-1.676 (0.817)	-3.709 (1.550)	-3.453 (1.009)	-2.682 (0.900)
Outcome Mean	83.31	73.68	82.22	89.61
English Course Grade	-1.719 (0.722)	-3.225 (1.448)	-2.727 (0.963)	-2.747 (0.902)
Outcome Mean	84.60	76.92	84.29	89.28
Math I EOC Score (<i>Standardized</i>)	0.040 (0.052)	0.001 (0.065)	-0.056 (0.058)	-0.017 (0.090)
Outcome Mean	0.003	-0.635	0.072	0.811
Days Absent	1.258 (0.441)	1.613 ^p (1.137)	2.119 ^p (0.607)	0.003 ^p (0.478)
Outcome Mean	5.728	7.798	6.197	4.119
Prob. Chronic Absence	0.025 (0.013)	0.039 ^p (0.034)	0.046 ^p (0.019)	-0.012 ^p (0.014)
Outcome Mean	0.045	0.089	0.053	0.012
Observations	254,832	80,072	91,451	83,309

NOTES: Estimates presented come from difference-in-discontinuity regressions that pool 9th grade academic years 2013-2014 through 2015-2016. “Analytic Sample” restricts to those with a valid math test score in grade 8. Ability levels are derived by within-year standardized 8th grade math performance. subscript *p* denotes effects that are statistically different across ability groups. Observation numbers refer to GPA and absence outcomes. Approximately 40% of 9th graders do not take the Math I EOC in 9th grade. For the analytic sample, we therefore have 151,429 observations in our Math I EOC regression. Heteroskedastic-robust standard errors are presented in parentheses. The outcome means displayed refer to the estimated constant in the corresponding regression specification.

by Math I EOC scores or numeric marks in math and English (which decline by 1.7 points each). Instead, increased grade leniency induced students to become 1.3 days more absent in school compared to the counterfactual, corresponding to a 22% increase in school absences. Finally, chronic absenteeism rates increased by 0.3 percentage points, indicating that the effect on school absences might be driven by students on the upper tail of the distribution that were, counterfactually, already exhibiting high levels of absenteeism.

In [Section II](#), changes in grading leniency descriptively revealed stark differences in responses among students based on their academic preparedness, indicating that average effects may hide important heterogeneity among student groups. We unpack the average treatment effect estimates by examining heterogeneity among low, medium, and high ability groups based on 8th grade Math

Figure 6: Simulated Gains in GPA and Median Performance, by Ability Decile



NOTES: We simulate mechanical gains in 9th grade GPA following the policy change for the subset of students with numeric grades on their transcript. These effects are shown in blue dots. We also display diff-in-disc estimates on student GPA that correspond to the estimates presented in Table 3 with the corresponding 95% confidence intervals.

end of grade (EOG) performance. We report effects for each student ability group in columns (2) through (4) of Table 3.

The point estimates for GPA are very similar across ability groups. It is somewhat surprising that the low ability group did not have larger GPA gains than the high ability group because the largest increases in leniency were at the bottom of the performance distribution. At the same time, we show in Figure 1 that teachers might be less willing to round at the pass threshold as a result of this policy. This could lower numeric grades and erode the effects differentially for students at the bottom of the distribution, which would explain the lack of precision of the GPA effect we recover for lower ability students.

In contrast to findings for GPA, we show that low and medium ability students are the main drivers of increased (chronic) absenteeism. Low and medium ability students see increases in absences of 1.6 and 2.1 days, while high ability students experience no detectable change. These effects are statistically different across ability groups.

It is puzzling to find that this policy both increased student GPA and absences, measures that have jointly been used to examine student effort and non-cognitive skills (Hastings, Neilson and Zimmerman, 2012; Jackson, 2018; Stinebrickner and Stinebrickner, 2008). However, our context is unique in that the policy was expected to mechanically increase GPA. Thus, to contextualize our estimates on GPA effects, it is helpful to first understand the extent to which this policy increases

GPA by design. To do so, we draw from pre-treatment (2015) student numeric grade data and calculate a *simulated* GPA using the new grading scale (2016). The difference between the actual and simulated GPA measures helps us understand how GPA would change in the absence of any endogenous response.¹⁷ Figure 6 reports simulated changes in GPA on average and by ability tercile (blue dots) and the corresponding diff-in-disc estimates for GPA (pink confidence interval bands). Overall, the average simulated GPA difference is estimated to be 0.252 points. Thus, *ex-ante*, we would have expected a mechanical increase in GPA of approximately 0.25 points by policy design. Instead, our diff-in-disc estimates indicate an increase of only 0.127 points, a 51% reduction. Furthermore, the highest mechanical gains were concentrated among low ability students.

V.C Longer-Run Effects

We provided evidence that grade leniency differentially impacted students of varying ability levels during the first year of the policy’s implementation. This section explores whether these differences compound over time. On one hand, the initial exposure to this policy could have caused gaps that subsequently disappeared as students adjusted to the new grading standard. On the other hand, gaps may widen each year if the effects of the policy generated lasting changes in the education trajectories of students. If the latter proves to be true, it would suggest that academic leniency and grade inflation can exacerbate achievement gaps and lessen human capital accumulation for students already at a deficit. We explore this by following our cohort of 9th graders for three additional years through high school completion.

In our setting, the estimation of longer-run outcomes is less straightforward than the short-run 9th grade results presented in Section V.B. To understand why, recall that the policy applied to all high school students at the time of enactment, regardless of grade. This means that as our first cohort of 9th graders progress through high school, our control cohorts will also become exposed to treatment.¹⁸ To overcome these concerns, we follow our 9th graders through high school by estimating diff-in-disc models that fix the control cohorts to always be those in the academic grade

¹⁷ The main assumption underlying this exercise is that in the absence of the new grading policy, post-treatment numeric grade distributions would have been similar to pre-treatment ones. While we cannot test for this directly, we provide evidence in Figure B5 that pre-reform numeric grade distributions were relatively fixed over time up until the year of the policy change. Note that pre-treatment numeric grade distributions are the product of both student effort and teacher grading practices given the state-mandated grade policy of the time. Thus, this simulation exercise recovers the change in GPA that would have been generated if both students and teachers had completely ignored the policy change and maintained their business-as-usual practices.

¹⁸ For example, as our initial 9th grade cohort transition to 10th grade, diff-in-disc estimates for 10th grade would compare them to our two control cohorts of 10th graders (2015 and 2016). One of the 10th grade cohorts experiences a year of treatment. Figure C1 illustrates how control exposure to treatment evolves over time.

of interest during 2014 and 2015. In other words, we compare 10th graders in 2017 to 10th graders in 2014 and 2015, 11th graders in 2018 to 11th graders in 2014 and 2015, and 12th graders in 2019 to 12th graders in 2014 and 2015.

Table 4: Dynamic Results for Main Analysis Cohort

		<i>Ability Level</i>		
	Analytic Sample (1)	Low (2)	Medium (3)	High (4)
<i>Panel A: 10th Grade</i>				
Core Academic GPA	−0.139 (0.048)	−0.184 (0.085)	−0.069 (0.065)	−0.102 (0.064)
Days Absent	1.576 (0.444)	2.761 (1.075)	1.447 (0.666)	0.424 (0.571)
Chronic Absence	0.026 (0.015)	0.056 (0.035)	0.033 (0.023)	−0.014 (0.018)
<i>Panel B: 11th Grade</i>				
Core Academic GPA	−0.157 (0.049)	−0.204 (0.091)	−0.117 (0.068)	−0.143 (0.069)
Days Absent	2.859 (0.541)	4.763 (1.321)	3.039 (0.823)	1.633 (0.733)
Chronic Absence	0.058 (0.018)	0.114 (0.043)	0.059 (0.028)	0.030 (0.023)
ACT	−0.498 (0.262)	−0.536 (0.299)	−0.517 (0.278)	−0.311 (0.423)
<i>Panel C: 12th Grade</i>				
Core Academic GPA	−0.048 (0.049)	−0.021 (0.092)	−0.161 (0.073)	−0.069 (0.074)
Days Absent	0.297 (0.657)	2.567 (1.540)	0.603 (1.049)	−0.405 (0.886)
Chronic Absence	0.017 (0.020)	0.080 (0.046)	−0.004 (0.032)	0.026 (0.028)
Graduation	0.033 (0.013)	0.045 (0.028)	0.053 (0.021)	0.012 (0.021)
College Intent	−0.066 (0.028)	−0.102 (0.048)	−0.106 (0.043)	−0.056 (0.043)

NOTES: Estimates presented come from difference-in-discontinuity regressions that follow cohorts of students over time. The first row tracks our main cohort (the treatment cohort randomly exposed to the 10-point scale in 9th grade). The next row considers instead the set of 10th graders whose birth date randomly assigned them to the first year of the 10-point scale in 10th grade. The next two rows follow this intuition for 11th and 12th graders. Each specification controls for gender, race, and SES level. Heteroskedastic-robust standard errors are presented in parentheses.

Table 4 displays longer-run difference-in-discontinuity results for our initial 9th grade cohort

throughout the remaining years of high school. As before, we present estimates for the full sample and by ability subgroups. Panels A, B, and C display results for 10th, 11th, and 12th grade, respectively. These findings indicate that the policy’s effects on increased absences persist and worsen throughout high school. In 11th grade, our treatment cohort is 2.9 days more absent in school compared to our 11th grade control cohorts, an effect that is more than twice as large as the one recovered for the first year of exposure. When exploring long-run effects across ability groups, our findings suggest that long-term increase in absences are driven by students in the middle and bottom of the ability distribution.

At the same time, increased grade leniency caused GPA to decline immediately following the 9th grade increase, with the recovered point estimate suggesting that this decline exceeded the gains made in the previous year. By 11th grade, students experience net losses to their GPA on average. These results support our interpretation that the initial gains in GPA were primarily mechanical in nature. Furthermore, they point to the possible unintended losses that occurred over time as a result of increased leniency. However, despite stark heterogeneous long-run effects on student absences across ability groups, we are unable to detect systematic differences in long-term GPA effects.

Given these differences in the effects of the policy on students’ behavioral response, we explore potential repercussions on college readiness, which we proxy using ACT scores, high school graduation, and college intent. The ACT score provides us with a nationally standardized measure of cognitive ability for all students in the state. Estimates presented in panel B of [Table 4](#) indicate a negative effect on ACT scores for the full sample that is statistically significant at the 10% level. These estimates show that early exposure to the 10-point scale caused a decrease in ACT scores of 0.5 points, or approximately 2.4%.¹⁹ Heterogeneous effects suggest greater declines in ACT performance for students at the bottom and middle of the ability distribution, although we cannot reject the null hypothesis that differences between groups are statistically different than zero.

Negative effects on ACT performance are accompanied by negative effects on college-going aspirations. We find that college intent declines by 6.6 p.p., an effect that seems to be driven by low and medium ability students. Despite this, we find a positive effect on the likelihood to graduate high school of 3.3 p.p. Again, low and medium ability students drive these results. This discrepancy between high school graduation and college readiness indicators aligns with national

¹⁹ This finding is in line with the literature on the effects of increased absences on students’ test score performance ([Liu, Lee and Gershenson, 2021](#)).

trends of increases in high school graduation that do not translate into similar increases in college enrollment.

In order to jointly interpret our short-term and longer-run results, it is helpful to consider three potential channels for benefits and costs of the policy. A key benefit is that students are more likely to have higher GPAs due to the reduction in standards. Through this change, it is easier to pass courses, which may make it easier to graduate from high school. However, there are also two potential costs that accompany this change. First, if students respond to more lenient standards with lower effort levels, they may experience learning loss. In our setting, we argue that this loss in learning and in human capital is captured by lowered ACT scores. Second, reduced effort, as observed here through increased absenteeism, has broader implications for human capital outside of learning loss.

With these three channels in mind, we first focus on low ability students. Despite being the group with the highest *expected* GPA gains, these students only see about a third of the expected gains in the first year of the policy (see [Figure 6](#)) and experience net losses in GPA thereafter. However, in the long-run results, we see significant increases in high school graduation rates. In terms of the costs associated with this policy, we find that the low ability group increased their absence rate significantly and consistently throughout high school. Ultimately, we find learning losses for this group through lower ACT scores and lower college-going intentions.

At the other end of the distribution, high ability students also experienced short term GPA gains that faded over time, but they did not experience gains in high school graduation rates. In contrast to their lower ability peers, we find weaker evidence that this group systematically reduced their effort across high school years. Furthermore, we are unable to detect statistically significant evidence of learning loss for this group, nor do we find significant evidence of reduced aspirations to pursue postsecondary education.

Taken together, we interpret this evidence as indicative of stark heterogeneous response by ability with benefits that fade out over time and costs that persist, especially among lower ability students. Ultimately, these results point to widening achievement gaps in secondary school.

V.D Older Cohorts

While the 9th grade cohort is the focus of this paper, we also explore the effects of the change in leniency on effort and other outcomes among older cohorts of students to better understand the broader policy implications of our results. [Table 5](#) presents short- and long-term diff-in-disc

Table 5: Dynamic Results for Older Cohorts

	Cohort at Time of Policy Change		
	10th Graders (1)	11th Graders (2)	12th Graders (3)
<i>Panel A: 10th Grade Outcomes</i>			
Core GPA	0.105 (0.049)		
Days Absent	0.647 (0.446)		
Chronic Absence	0.020 (0.015)		
<i>Panel B: 11th Grade Outcomes</i>			
Core GPA	-0.032 (0.050)	0.118 (0.051)	
Days Absent	0.814 (0.518)	1.221 (0.515)	
Chronic Absence	0.017 (0.017)	0.020 (0.017)	
ACT	0.114 (0.269)	0.091 (0.276)	
<i>Panel C: 12th Grade Outcomes</i>			
Core GPA	-0.082 (0.050)	-0.044 (0.053)	0.201 (0.052)
Days Absent	1.530 (0.653)	0.333 (0.620)	0.331 (0.597)
Chronic Absence	0.032 (0.020)	-0.024 (0.020)	0.009 (0.019)
Graduation	-0.011 (0.015)	0.015 (0.015)	-0.008 (0.015)
College Intent	-0.062 (0.029)	-0.024 (0.030)	-0.031 (0.030)

NOTES: Estimates presented come from difference-in-discontinuity regressions that pool 2012-2013, 2013-2014, and 2014-2015 as control years and the corresponding years when the given focus cohort was in the stated grade as treatment years. For instance, the first entry for 11th grade GPA is the corresponding diff-in-disc estimate for the 10th graders whose birth date randomly assigned them to the first year of the 10-point policy *in 10th grade*. Core academic GPA's are grade specific and not cumulative. Graduation refers to the probability to graduate in 12th grade. College intent is measured using a student survey administered in 12th grade and refers to the probability to indicate interest in attending a 4-year university. Heteroskedastic-robust standard errors are presented in parentheses.

estimates for these cohorts of students. These results are consistent, albeit somewhat attenuated, with our main findings.

To understand the immediate effects of the policy for each grade level, we focus on the “diagonal”

portion of [Table 5](#): the first column in Panel A, the second column in Panel B, and the third column in Panel C. These results demonstrate that the policy mechanically increased student GPA across all grades by 0.1-0.2 points. Absences increase as well, although these estimates are noisily estimated for 10th and 12th graders. The somewhat attenuated effect on absences for these older cohort of students might be indicative of the pivotal role that 9th grade can have on establishing norms in high school. Additionally, we do not find that the policy had an immediate effect on 11th graders’ ACT performance or 12th graders’ propensity to graduate high school or indicate college intention. These null effects point to the fact that the long-term effects shown in [Section V.C](#) for our main cohort of interest are the result of continued exposure to the policy.

In order to understand the compounding effects of policy exposure for older students, we focus on the “off-diagonal” parts of [Table 5](#). In line with our main results, these estimates show that GPA effects fade over time for older cohorts while absence effects persist. However, when examining effects on college readiness proxies, we find some differences relative to our main findings. First, the policy did not change ACT scores for either 10th or 11th graders. Second, negative effects on college intent are only present in our 10th grade cohort. Finally, we detect no graduation effects for either of these cohorts. Again, these results suggest that negative effects on longer-term outcomes are related to more years of exposure to the lenient grading scheme and continue to point to the importance of 9th grade over later years in determining student effort.

VI Robustness Checks

Our main results are robust to a battery of alternative specifications, including the clustering of standard errors, changing the birth date bandwidth from 180 days to 90 days, and using a linear fit on a 60-day bandwidth, which is commonly used in the literature. Estimates for each of these checks are reported in panels B through D of [Table A2](#). We present results for three of our short-run outcomes: GPA, math grades, and number of days absent. We do not find any meaningful changes in our results under these competing specifications. We also perform five additional robustness checks that we discuss below.

VI.A Placebo Analysis

We first conduct a placebo analysis in the style of [Bertrand, Mogstad and Mountjoy \(2021\)](#) to test for the reliability of our findings. This exercise estimates diff-in-disc regressions on alternative

birth date cutoffs and shows that no other birth date cutoff produces similar findings to the ones we present in this paper. For this, we construct a placebo cut point set to k additional days after the true birthday cut-off of October 17th (normalized to 0). We then develop viable “control” and “treatment” birthday windows that satisfy the following properties: first, the windows cannot overlap with one another; second, the windows cannot contain the true cut point of $b_i = 0$; third, the windows should not be substantially influenced by noncompliance close to the true cut point. In line with this third property, we exclude from the placebo analysis a 30-day buffer on either side of the true cut-off.

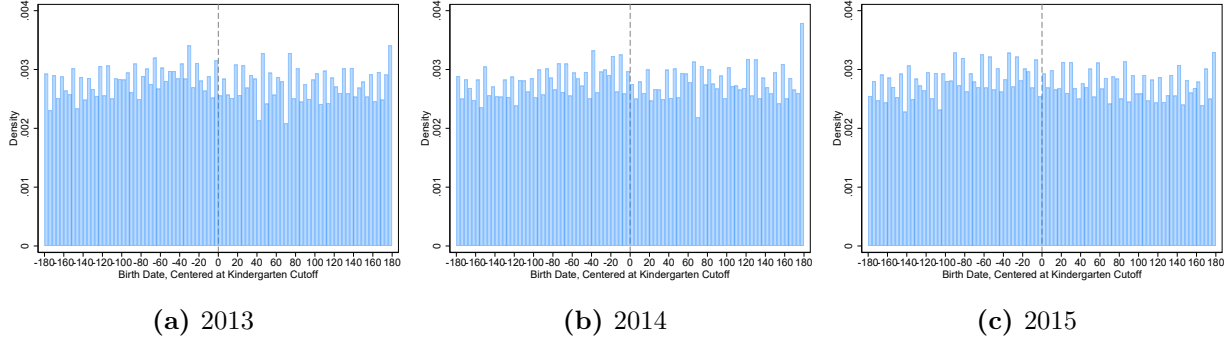
We consider nine discrete values for the placebo cut-off, $k = \{120, 135, \dots, 240\}$. For $k < 180$, we include in the birthday windows any student with $b_i \in (30, 150 + k)$. Similarly, any $k \geq 180$ generates a set of birthday windows comprised of students with $b_i \in (k - 150, 330)$. This approach allows us to maintain at least 90 days on either side of the cut-off for any specification. Thus, for each k , we derive a k -specific set of control and treatment groups to conduct our analysis on. For consistency, we pool academic years 2013-2014 and 2014-2015 to create the control window and use academic year 2015-2016 as the treatment window. Due to the fact that these cut points are, by design, unrelated to the true kindergarten entry rule, we cannot conduct a *fuzzy* difference-in-discontinuity design. Instead, we utilize the sharp design to test for whether or not our main results in 9th grade can be replicated.

Figure B6 displays the results of our placebo analysis. Each panel focuses on one of our five short-run outcomes of interest: 9th grade GPA, math course averages, standardized Math I EOC scores, number of days absent, and the likelihood to be chronically absent. We report sharp difference-in-discontinuity estimates for each outcome and for each ability tercile, each with corresponding robust standard errors. Across the nine placebo cut points and 180 combinations of outcomes, we recover only one statistically significant estimate. Importantly, the point estimates do not match those found in Section V, which points to a strong reliability in our main specification and results.

VI.B Manipulation of the Running Variable

The validity of a regression discontinuity design fails whenever agents can plausibly manipulate the running variable and their likelihood of receiving treatment (McCrary, 2008). In our setting, this would require that parents not only strategically induce the birth of their children to coincide with the October 17th start date cutoff, but that they do so distinctly in the year 2000 compared to 1999 and 1998, as they anticipate the policy change. Given that the policy was announced

Figure 7: Distribution of Birth Dates Around the Cut Point



NOTES: For visual ease, we bin birth dates in groups of four. For each cohort window, we plot 180 days on either side of the cutoff and display 45 bars in the figure above on both the left and right of 0.

approximately 15 years after this cohort of parents would be making fertility decisions, we have no reason to believe this is the case. Furthermore, [Dickert-Conlin and Elder \(2010\)](#) find no evidence of school-age start date manipulation via induction or C-sections using a national dataset of birth records. In fact, even when parents have a sizable financial incentive to manipulate the timing of their child’s birth, [LaLumia, Sallee and Turner \(2015\)](#) show that these effects are small.

Still, we formally test for manipulation using the procedure outlined in [Cattaneo, Jansson and Ma \(2018\)](#), which estimates a local polynomial density on either side of the cutoff and tests a null hypothesis that the limit of both functions approaching from either side of the cutoff are equal. Using an optimally-selected 51-day bandwidth and pooled cohort running variable data, the test yields an associated p -value of 0.68. Thus, we fail to reject the null hypothesis that the density functions are equal at the cutoff and find no evidence of systematic manipulation of the running variable. Visual evidence presented in [Figure 7](#) shows a smooth distribution of birthdays around the kindergarten entry cutoff for all cohorts, which further supports this finding. With these results in mind, we proceed under the assumption that birth dates exogenously sort students to a kindergarten entry year and are not manipulated by parents.

VI.C Covariate Smoothness

We next test for covariate smoothness at the threshold and policy-relevant window. Our empirical analysis functions as a quasi-experiment only if observable attributes of students at our policy-relevant cohort window trend smoothly at the boundary of kindergarten entry assignment in comparison to the non-policy-relevant cohorts. As an example, our assumption of random as-

Table 6: Fuzzy Difference-in-Discontinuity Estimates for Covariates

Female (1)	URM (2)	EDS (3)	Rurality (4)	LEP (5)	Grade 8 EOG (6)
0.001 (0.027)	-0.038 (0.026)	0.034 (0.027)	0.010 (0.027)	0.007 (0.010)	-0.039 (0.053)

NOTES: Estimates presented come from difference-in-discontinuity regressions that pool 9th grade academic years 2013-2014 through 2015-2016 and estimated via Equation 1. We present estimates for τ in our main specification without covariates. Heteroskedastic-robust standard errors are presented in parentheses.

signment fails if there is a change in the rate at which economically disadvantaged students comply with the assignment mechanism at our policy-relevant cohort window. If this were the case, our difference-in-discontinuity estimate would additionally capture socioeconomic effects, biasing and eliminating any causal interpretation of our results.

Table 6 presents the difference-in-discontinuity estimates, estimated via Equation 1, for student demographics (gender, race, EDS, rurality, and Limited English Proficiency status (LEP)) and student pre-treatment academic preparedness (8th grade Math EOG).²⁰ Overall, we find no evidence to suggest differential discontinuity around the threshold between the policy-relevant and non-policy-relevant window cohorts. We take these results to suggest that our assignment mechanism operates identically for different groups of students.

VI.D Selection of Control Cohort Window

In this exercise we show that our main results are robust to the choice of cohorts that we include to form our control windows. First, in Panel B of Table A3, we expand the control group to include two post-period control windows that span from the 2016-2017 academic year to the 2017-2018 one. Our preferred specification does not include these cohorts because of potential treatment contamination. While not required, the North Carolina grading policy also allowed middle schools to adopt the new 10-point grading scale to align expectations across school levels. Middle school take-up was therefore staggered and non-universal. Given that transcript data is available only for high school grades, we are unable to identify which middle schools adopted the policy, nor when. However, based on publicly available district records, only one district (out of 115) in the state maintained the 7-point scale for middle schools through 2024. Given that we cannot exclude from the analysis early middle school adopters, we cannot ensure that students on either side of the birth

²⁰ Figure B7 and Figure B8 present analogous visual evidence using a quadratic fit over a binscatter of the relevant covariates across a 180-day bandwidth.

date cutoff in post-period comparisons have the same amount of exposure to the new grading scale. This makes post-period comparisons unsuitable for the estimation of age effects. Even so, in Panel B of [Table A3](#) we include a fully pooled version of the model. Qualitatively, these estimates point to the same interpretation as our main results.

We next vary which pre-reform cohorts are included in the control. Our preferred specification does not include students whose birthday would randomly assign them to begin high school in the 2011-2012 academic year because these students were the last to experience a statewide regime prior to the adoption of Common Core. Still, we show in Panel C of [Table A2](#) that our results are not substantively changed by including this academic year in our control group. Nearly every point estimate is meaningfully unchanged, except for the effect on absences, which attenuates but remains significant. In Panel D of [Table A2](#), we instead tighten the control window to include only the most recent preceding cohort of students. As before, the estimates on our main outcomes of interest remain largely unchanged.

VI.E Changes in Teacher Quality

Our final robustness check explores the role of changes in important education inputs that can potentially bias our student-level estimates. We focus on one of the most important inputs to the production of education, teacher quality ([Chetty, Friedman and Rockoff, 2014](#)). A change in teacher quality at the same time as the introduction of the policy would distort the estimated effects associated with academic leniency. For this exercise, we measure teacher quality by value-added (VA).

We link students in our analytical sample that take Math I in 9th grade (subsample of students with a valid Math I EOC score) to Math I teachers using course membership files. NCERDC course membership files contain records of the teacher who taught a particular course and the students that took that course in any given school year. This allows for a straightforward match of students to teachers.²¹ Overall, we are able to match 81% of the students in our analytical sample to a teacher record.

In a similar fashion to [Jackson \(2018\)](#), we consider a 9th grader i who takes math with teacher j in classroom c at school s in year t . We consider Math I EOC scores as our outcome of interest,

²¹ For more information see NCERDC Technical Report 1B and Technical Report #5.

Table 7: Fuzzy Difference-in-Discontinuity Estimates for Teacher Value-Added

	Analytic Sample (1)	Low Ability (2)	Medium Ability (3)	High Ability (4)
Empirical Bayes Leave-One-Out	0.002 (0.004)	-0.001 (0.006)	0.001 (0.005)	0.008 (0.009)
Outcome Mean	0.008	0.005	0.009	0.009
Leave-One-Out	0.001 (0.007)	-0.005 (0.012)	0.001 (0.010)	0.009 (0.017)
Outcome Mean	0.010	0.004	0.011	0.013
Observations	132,810	55,241	56,355	21,214

NOTES: Estimates presented come from difference-in-discontinuity regressions that pool 9th grade academic years 2012-2013 through 2015-2016. Empirical Bayes estimates are constructed using the procedure outlined in [Jackson \(2018\)](#). Leave-One-Out estimates refer to value-added estimates that do not scale by reliability. “Analytic Sample” restricts to those with a valid math test score in grade 8. Ability levels are derived from within-year standardized 8th grade math performance. Each specification controls for gender, race, and SES level. Heteroskedastic-robust standard errors are presented in parentheses.

denoted by y_{ijcst} , and estimate equations of the following form:

$$y_{ijcst} = \psi y_{i,t-1} + \boldsymbol{\zeta}' \mathbf{x}_{icst} + \alpha_{st} + e_{ijcst}. \quad (2)$$

In the above, $y_{i,t-1}$ is a lagged measure of ability (8th grade math EOG score), $\mathbf{x}_{icst}$ is a vector of student controls that includes race, gender, socioeconomic status, and class size, α_{st} is a school-by-year fixed effect, and e_{ijcst} is the student-level residual. This residual is assumed to have a decomposition that follows

$$e_{ijcst} = \theta_j + \varepsilon_{jcst} + \varepsilon_{ijcst}, \quad (3)$$

where θ_j is the VA of teacher j and $\{\varepsilon_{jcst}, \varepsilon_{ijcst}\}$ are random shocks at the class and student level, respectively. We estimate [Equation 2](#) using 9th grade academic years 2012-2013 through 2017-2018 to include more information on teachers. The balance test follows our main analysis and uses only academic years 2013-2014 through 2015-2016.

We estimate leave-one-out VA estimates by predicting the effect that a teacher has on Math I scores using the *other* years of data that include that teacher. For example, we estimate VA for 9th graders in the 2015-2016 academic year using the predicted VA for that teacher in the three preceding years and the two following years. This estimate, denoted by $\hat{\theta}_{j,-t} = \bar{e}_{j,-t}$, is then used to construct leave-one-out Empirical Bayes (EB) estimates of VA. The EB VA scales the original VA by a measure of reliability, λ_j , which follows directly from [Jackson \(2018\)](#).²² For completeness,

²² For finer details on the construction of the EB estimates, please see [Jackson \(2018\)](#).

we write the EB estimate as $\hat{\mu}_{jt} = \lambda_j \hat{\theta}_{j,-t}$.

As shown in [Table 7](#), we find no significant effect at the cut point when treating $\hat{\mu}_{jt}$ as the outcome variable in our main estimating equation. We run this separately for each ability level and find no effect at any of these margins. This evidence provides assurance that our main results are not driven by changes in the exposure to effective teachers. Additionally, these findings are robust to the use of $\hat{\theta}_{j,-t}$ as our measure of VA.

VII Conclusion

The economic benefits of high school graduation are large and persistent, which has led local governments across the country to enact policy with the explicit goal of increasing graduation rates. One popular mechanism designed to boost the performance of low-achieving students is to relax stringent academic standards to help students meet graduation requirements and obtain a high school degree. As a result, the last decade has seen steady increases in graduation rates and grade point averages in U.S. high schools without any measured increase in student achievement. At face value, a decline in academic standards can naturally lead to more equitable outcomes since a larger share of students meet graduation requirements and obtain a high school diploma. However, this ignores the possible role that academic standards have on students' effort, which directly impacts their accumulation of human capital.

We show that lowered grading standards have the potential to exacerbate achievement gaps. Under a fuzzy difference-in-discontinuity research design that leverages variation from statewide grading policy and school entry rules, we find that increased academic leniency led to mechanical GPA gains and large effort reductions. Reductions in effort, as measured by attendance, are especially concentrated among students with low and medium ability demonstrated by prior performance. These heterogeneous effects compound over time and lead to widening gaps in effort, achievement (ACT scores), and college-going intention. Therefore, the short-run gains of artificially raising GPA scores may be detrimental, especially when lowered standards are not associated with increases to human capital accumulation. This paper highlights the importance of understanding how students respond to changes in policy, and how policy designed without these responses in mind may yield large, unintended consequences.

References

- Ahn, Tom, Peter Arcidiacono, Amy Hopson, and James Thomas. 2024. “Equilibrium grading policies with implications for female interest in STEM courses.” *Econometrica*, 92(3): 849–880.
- Allensworth, Elaine M, and John Q Easton. 2007. “What Matters for Staying On-Track and Graduating in Chicago Public High Schools: A Close Look at Course Grades, Failures, and Attendance in the Freshman Year. Research Report.” *Consortium on Chicago School Research*.
- Angrist, Joshua D, and Alan B Krueger. 1991. “Does compulsory school attendance affect schooling and earnings?” *The Quarterly Journal of Economics*, 106(4): 979–1014.
- Asker, Erdal. 2024. “The Effect of Compulsory Education on Age at First Marriage and Teenage Fertility: Evidence from a Difference-in-Discontinuity Design.” *Working Paper*.
- Babcock, Philip. 2010. “Real costs of nominal grade inflation? New evidence from student course evaluations.” *Economic inquiry*, 48(4): 983–996.
- Bar, Talia, Vrinda Kadiyali, and Asaf Zussman. 2009. “Grade information and grade inflation: The Cornell experiment.” *Journal of Economic Perspectives*, 23(3): 93–108.
- Behrman, Jere R, Ricardo Gomez-Carrera, Susan W Parker, Petra E Todd, and Weilong Zhang. 2024. “Starting Strong: Medium-and Longer-run Benefits of Mexico’s Universal Preschool Mandate.” *Unpublished manuscript*.
- Bertrand, Marianne, Magne Mogstad, and Jack Mountjoy. 2021. “Improving educational pathways to social mobility: evidence from Norway’s reform 94.” *Journal of Labor Economics*, 39(4): 965–1010.
- Betts, Julian R. 1998. “The impact of educational standards on the level and distribution of earnings.” *The American Economic Review*, 88(1): 266–275.
- Betts, Julian R, and Jeff Grogger. 2003. “The impact of grading standards on student achievement, educational attainment, and entry-level earnings.” *Economics of Education Review*, 22(4): 343–352.
- Blagg, Kristin, and Matthew Chingos. 2016. “Varsity Blues: Are High School Students Being Left Behind?”
- Butcher, Kristin F, Patrick J McEwan, and Akila Weerapana. 2014. “The Effects of an Anti-Grade Inflation Policy at Wellesley College.” *Journal of Economic Perspectives*, 28(3): 189–204.
- Butcher, Kristin, Patrick J McEwan, and Akila Weerapana. 2024. “Making the (letter) grade: The incentive effects of mandatory pass/fail courses.” *Education Finance and Policy*, 19(3): 385–408.

- Cattaneo, Matias D, Michael Jansson, and Xinwei Ma. 2018. "Manipulation testing based on density discontinuity." *The Stata Journal*, 18(1): 234–261.
- Chan, William, Li Hao, and Wing Suen. 2007. "A signaling theory of grade inflation." *International Economic Review*, 48(3): 1065–1090.
- Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American economic review*, 104(9): 2633–2679.
- Cook, Philip J, and Songman Kang. 2016. "Birthdays, schooling, and crime: Regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation." *American Economic Journal: Applied Economics*, 8(1): 33–57.
- Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." *Advances in Applied Microeconomics*, 14(1): 1–34.
- Dee, Thomas S, and Hans Henrik Sievertsen. 2018. "The gift of time? School starting age and mental health." *Health economics*, 27(5): 781–802.
- Dee, Thomas S, Will Dobbie, Brian A Jacob, and Jonah Rockoff. 2016. "The causes and consequences of test score manipulation: Evidence from the new york regents examinations." National Bureau of Economic Research.
- Deming, David, and Susan Dynarski. 2008. "The lengthening of childhood." *Journal of economic perspectives*, 22(3): 71–92.
- Denning, Jeffrey T, Eric R Eide, Kevin J Mumford, Richard W Patterson, and Merrill Warnick. 2022. "Why have college completion rates increased?" *American Economic Journal: Applied Economics*, 14(3): 1–29.
- Diamond, Rebecca, and Petra Persson. 2016. "The long-term consequences of teacher discretion in grading of high-stakes tests." National Bureau of Economic Research.
- Dickert-Conlin, Stacy, and Todd Elder. 2010. "Suburban legend: School cutoff dates and the timing of births." *Economics of Education Review*, 29(5): 826–841.
- Durden, Garey C, and Larry V Ellis. 1995. "The effects of attendance on student learning in principles of economics." *The American Economic Review*, 85(2): 343–346.
- Elder, Todd E. 2010. "The importance of relative standards in ADHD diagnoses: evidence based on exact birth dates." *Journal of health economics*, 29(5): 641–656.

- Espenshade, Thomas J, Lauren E Hale, and Chang Y Chung. 2005. "The frog pond revisited: High school academic context, class rank, and elite college admission." *Sociology of Education*, 78(4): 269–293.
- Figlio, David N. 2006. "Testing, Crime and Punishment." *Journal of Public Economics*, 90(4-5): 837–851.
- Figlio, David N, and Maurice E Lucas. 2004. "Do high grading standards affect student performance?" *Journal of Public Economics*, 88(9-10): 1815–1834.
- Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics*, 2(3): 205–27.
- Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano. 2016. "Do fiscal rules matter?" *American Economic Journal: Applied Economics*, 1–30.
- Hansen, Anne Toft, Ulrik Hvidman, and Hans Henrik Sievertsen. 2023. "Grades and employer learning." *Journal of Labor Economics*.
- Harris, Douglas N, Lihan Liu, Nathan Barrett, and Ruoxi Li. 2023. "Is the rise in high school graduation rates real? High-stakes school accountability and strategic behavior." *Labour Economics*, 82: 102355.
- Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman. 2012. "The effect of school choice on intrinsic motivation and academic outcomes." National Bureau of Economic Research Working Paper 18324.
- Heckman, James J. 2006. "Skill formation and the economics of investing in disadvantaged children." *Science*, 312(5782): 1900–1902.
- Hurwitz, Michael, and Jason Lee. 2018. "Grade inflation and the role of standardized testing." *Measuring success: Testing, grades, and the future of college admissions*, 64–93.
- Hvidman, Ulrik, and Hans Henrik Sievertsen. 2021. "High-stakes grades and student behavior." *Journal of Human Resources*, 56(3): 821–849.
- Jackson, C Kirabo. 2018. "What do test scores miss? The importance of teacher effects on non-test score outcomes." *Journal of Political Economy*, 126(5): 2072–2107.
- Jacob, Brian A. 2005. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics*, 89(5-6): 761–796.
- Koretz, Daniel M. 2002. "Limitations in the Use of Achievement Tests as Measures of Educators' Productivity." *Journal of Human Resources*, 37(4): 752–777.

- LaLumia, Sara, James M Sallee, and Nicholas Turner. 2015. “New evidence on taxes and the timing of birth.” *American Economic Journal: Economic Policy*, 7(2): 258–293.
- Landaud, Fanny, Éric Maurin, Barton Willage, and Alexander Willén. 2024. “The value of a high school gpa.” *Review of Economics and Statistics*, 1–24.
- Liu, Jing, Monica Lee, and Seth Gershenson. 2021. “The short-and long-run impacts of secondary school absences.” *Journal of Public Economics*, 199: 104441.
- Mac Iver, Martha Abele, Marc L Stein, Marcia H Davis, Robert W Balfanz, and Joanna Hornig Fox. 2019. “An efficacy study of a ninth-grade early warning indicator intervention.” *Journal of Research on Educational Effectiveness*, 12(3): 363–390.
- McCrary, Justin. 2008. “Manipulation of the running variable in the regression discontinuity design: A density test.” *Journal of Econometrics*, 142(2): 698–714.
- Metcalfe, Robert, Simon Burgess, and Steven Proud. 2019. “Students’ Effort and Educational Achievement: Using the Timing of the World Cup to Vary the Value of Leisure.” *Journal of Public Economics*, 172: 111–126.
- Minaya, Veronica. 2020. “Do differential grading standards across fields matter for major choice? Evidence from a policy change in Florida.” *Research in Higher Education*, 61(8): 943–965.
- Murnane, Richard J. 2013. “US high school graduation rates: Patterns and explanations.” *Journal of Economic Literature*, 51(2): 370–422.
- Navarro-Palau, Patricia. 2017. “Effects of differentiated school vouchers: Evidence from a policy change and date of birth cutoffs.” *Economics of Education Review*, 58: 86–107.
- Nordin, Martin, Gawain Heckley, and Ulf Gerdtham. 2019. “The impact of grade inflation on higher education enrolment and earnings.” *Economics of Education Review*, 73: 101936.
- Ordine, Patrizia, Giuseppe Rose, and Daniela Sposato. 2018. “Parents Know Them Better: The Effect of Optional Early Entry on Pupils’ Schooling Attainment.” *Economic Inquiry*, 56(3): 1678–1705.
- Persson, Petra, Xinyao Qiu, and Maya Rossin-Slater. 2021. “Family spillover effects of marginal diagnoses: The case of ADHD.” National Bureau of Economic Research.
- Słoczyński, Tymon, S Derya Uysal, and Jeffrey M Wooldridge. 2024. “Abadie’s Kappa and Weighting Estimators of the Local Average Treatment Effect.” *Journal of Business & Economic Statistics*, 1–14.
- Stinebrickner, Ralph, and Todd R. Stinebrickner. 2008. “The Causal Effect of Studying on Academic

- Performance.” *The B.E. Journal of Economic Analysis & Policy*, 8(1).
- Todd, Petra E, and Kenneth I Wolpin. 2003. “On the specification and estimation of the production function for cognitive achievement.” *The Economic Journal*, 113(485): F3–F33.

Supplemental Appendix for “The Unintended Consequences of Academic Leniency” by A. Brooks Bowden, Viviana Rodriguez, and Zach Weingarten

A Additional Tables

Table A1: Mean Attributes of Compliers

	Female (1)	URM (2)	EDS (3)	Rurality (4)	Grade 8 EOG (5)
Compliers	0.550	0.428	0.420	0.485	0.189
Full Sample	0.518	0.437	0.444	0.486	0.113

NOTES: The above table displays mean characteristics of the designated group of students. The first row estimates these attributes for compliers using the methods in [Śloczyński, Uysal and Wooldridge \(2024\)](#). The second row presents raw averages of each attribute for the set of students born around the cut-off in the year 2000, which quasi-randomly assigns them to enter 9th grade in either 2015 or 2016. We restrict to those with non-missing covariate values, yielding a sample of 80,394 students.

Table A2: Alternative Specifications (Part I)

	Analytic Sample (1)	Ability Level		
		Low (2)	Medium (3)	High (4)
Panel A: Main Results (Reproduced)				
Core Academic GPA	0.127 (0.048)	0.127 (0.084)	0.112 (0.062)	0.104 (0.057)
Math Course Grade	-1.676 (0.817)	-3.709 (1.550)	-3.453 (1.009)	-2.682 (0.900)
Days Absent	1.261 (0.441)	1.613 (1.137)	2.125 (0.607)	0.003 (0.478)
Panel B: Clustered Standard Errors				
Core Academic GPA	0.127 (0.051)	0.127 (0.082)	0.112 (0.077)	0.104 (0.060)
Math Course Grade	-1.676 (0.740)	-3.709 (1.465)	-3.453 (1.110)	-2.682 (0.834)
Days Absent	1.261 (0.505)	1.613 (1.250)	2.125 (0.599)	0.003 (0.533)
Panel C: 90-Day Bandwidth (Quadratic)				
Core Academic GPA	0.113 (0.069)	0.141 (0.117)	0.196 (0.086)	-0.002 (0.087)
Math Course Grade	-2.310 (1.207)	-4.937 (2.224)	-2.611 (1.452)	-3.996 (1.443)
Days Absent	1.287 (0.630)	2.299 (1.615)	1.111 (0.827)	0.587 (0.731)
Panel D: 60-Day Bandwidth (Linear)				
Core Academic GPA	0.141 (0.057)	0.151 (0.098)	0.162 (0.072)	0.094 (0.069)
Math Course Grade	-1.734 (0.976)	-3.658 (1.839)	-3.045 (1.191)	-2.894 (1.098)
Days Absent	1.078 (0.518)	1.995 (1.338)	1.270 (0.697)	0.118 (0.568)

NOTES: Estimates presented come from difference-in-discontinuity regressions with the specification reported in the panel heading. Panel A reproduces our main results. Panel B reports results with standard errors clustered at the school-level. Panel C estimates the main specification using a quadratic fit and 90-day bandwidth. Panel D uses a linear specification on a 60-day bandwidth.

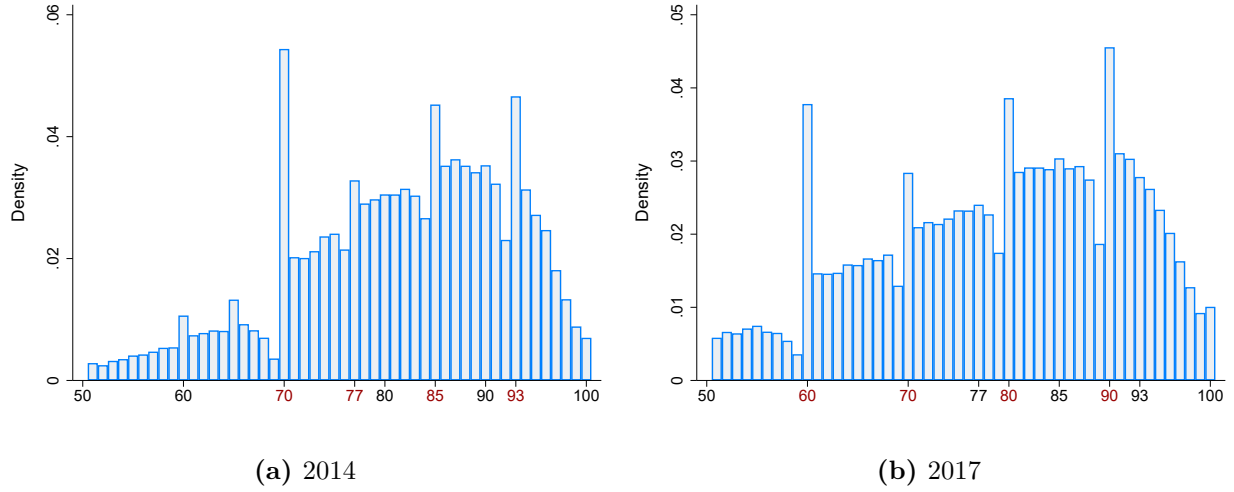
Table A3: Alternative Specifications (Part II)

		Ability Level		
	Analytic Sample	Low	Medium	High
	(1)	(2)	(3)	(4)
Panel A: Main Results (Reproduced)				
Core Academic GPA	0.127 (0.048)	0.127 (0.084)	0.112 (0.062)	0.104 (0.057)
Math Course Grade	-1.676 (0.817)	-3.709 (1.550)	-3.453 (1.009)	-2.682 (0.900)
Days Absent	1.261 (0.441)	1.613 (1.137)	2.125 (0.607)	0.003 (0.478)
Panel B: Post-Policy Pooled Control				
Core Academic GPA	0.107 (0.044)	0.105 (0.076)	0.082 (0.057)	0.056 (0.051)
Math Course Grade	-2.032 (0.776)	-4.749 (1.448)	-3.601 (0.961)	-2.790 (0.828)
Days Absent	0.614 (0.433)	0.743 (1.118)	1.238 (0.587)	-0.037 (0.451)
Panel C: Larger Pre-Period Control				
Core Academic GPA	0.137 (0.045)	0.123 (0.078)	0.174 (0.059)	0.112 (0.054)
Math Course Grade	-1.458 (0.774)	-3.258 (1.451)	-2.887 (0.963)	-2.313 (0.862)
Days Absent	0.815 (0.418)	1.000 (1.074)	1.456 (0.579)	-0.290 (0.455)
Panel D: Smaller Pre-Period Control				
Core Academic GPA	0.109 (0.056)	0.069 (0.097)	0.081 (0.071)	0.141 (0.069)
Math Course Grade	-1.197 (0.924)	-3.070 (1.808)	-3.502 (1.108)	-2.891 (1.015)
Days Absent	1.305 (0.503)	1.210 (1.307)	2.551 (0.657)	0.078 (0.567)

NOTES: Estimates presented come from difference-in-discontinuity regressions with the specification reported in the panel heading. Panel A reproduces our main results. Panel B presents estimates with the inclusion of two post-period control window (2016-2017 and 2017-2018). Panel C displays the estimates on a sample that includes the 2012-2013 academic year in addition to the other control years. Panel D includes only the 2014-2015 academic year as a control.

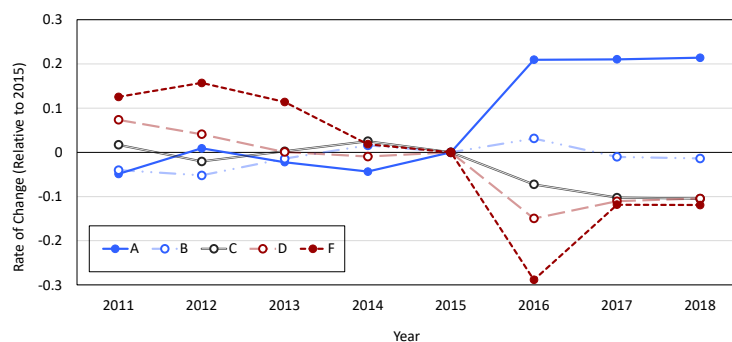
B Additional Figures

Figure B1: Additional Distributions of Final Course Averages in 9th Grade Math

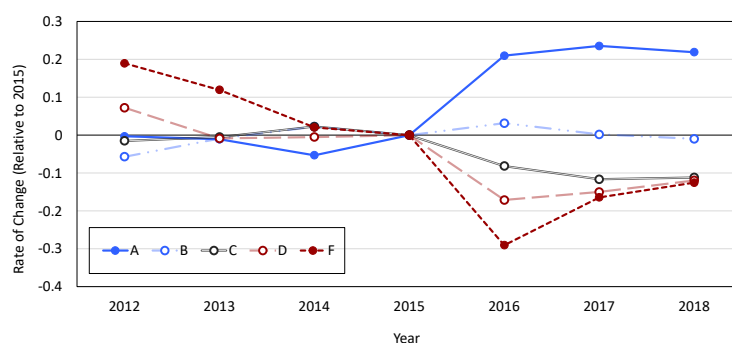


NOTES: As before, red labels denote cut grades. We fix both distributions to display only final grades at or above 50, accounting for the vast majority of transcript grades.

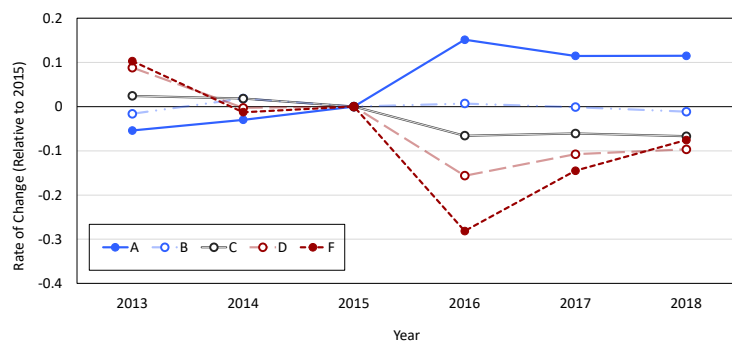
Figure B2: Change in Older Grade Letter Grade Shares, by Year



(a) 10th Grade



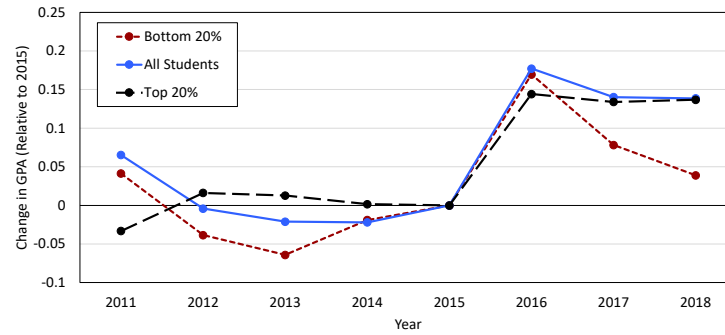
(b) 11th Grade



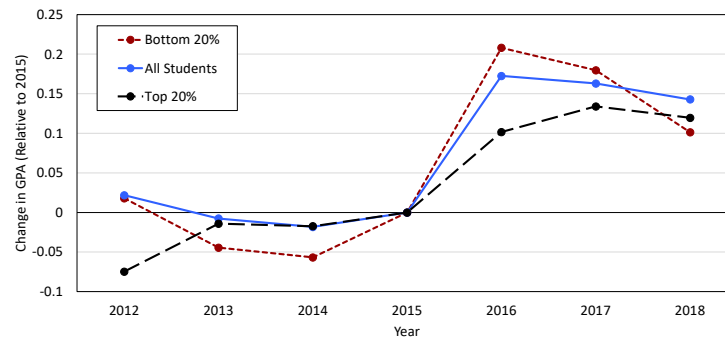
(c) 12th Grade

NOTES: Each panel plots the percent change in the share of each listed letter grade in core academic courses for the stated grade level, normalized against the base year of 2015. We restrict to those in our birth date sample and those with a valid 8th grade math score on record. This increasingly restricts older grades. For instance, we do not have information on 12th graders in 2011 when they were 8th graders (2007).

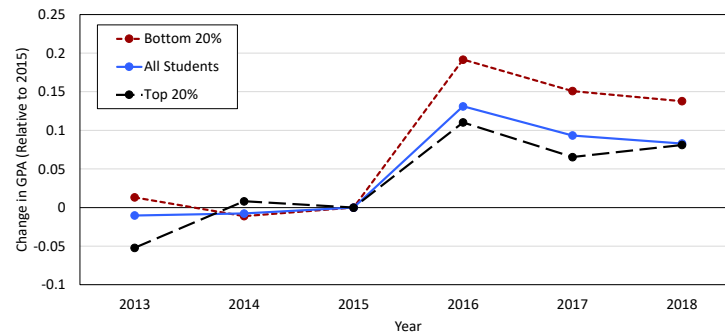
Figure B3: Change in Older Grade GPAs, by Year



(a) 10th Grade



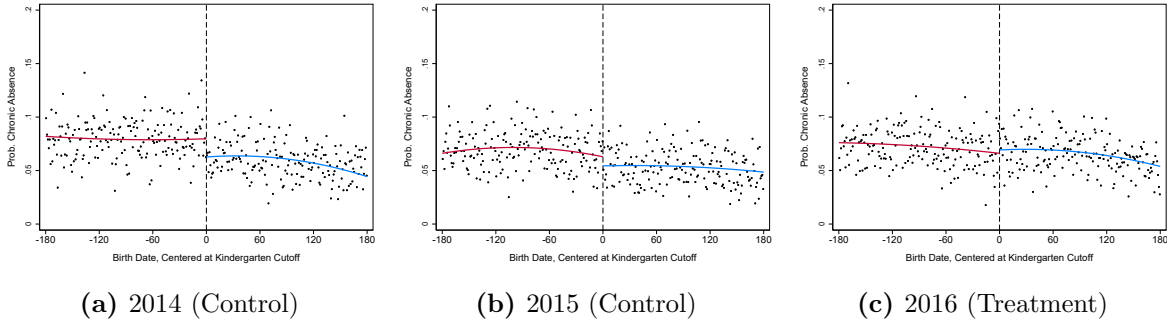
(b) 11th Grade



(c) 12th Grade

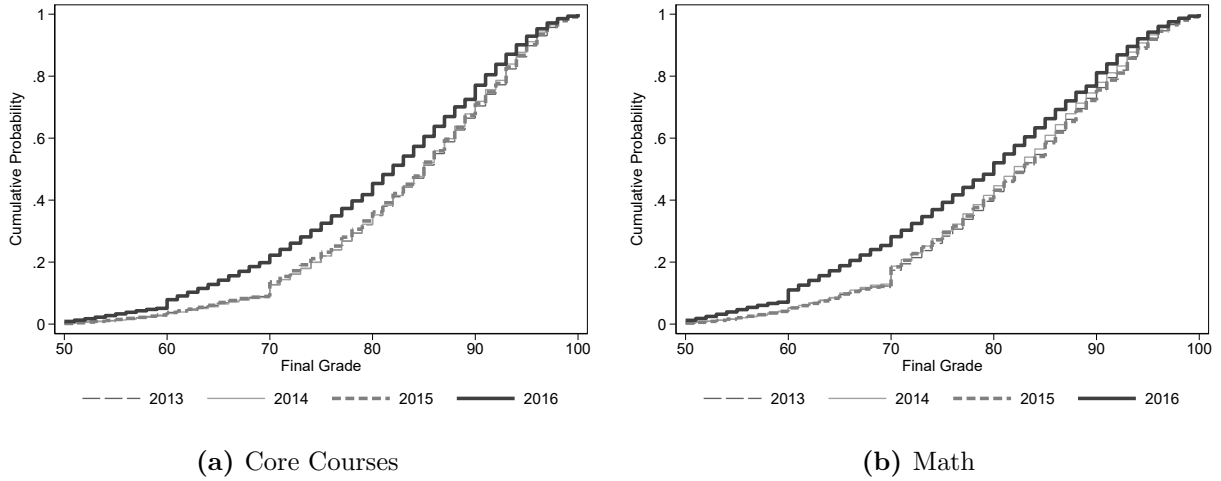
NOTES: Each panel plots the percent change in core academic GPA for the stated grade level, normalized against the base year of 2015. We restrict to those in our birth date sample and those with a valid 8th grade math score on record. This increasingly restricts older grades. For instance, we do not have information on 12th graders in 2011 when they were 8th graders (2007).

Figure B4: Regression Discontinuity of Chronic Absences Across Academic Years



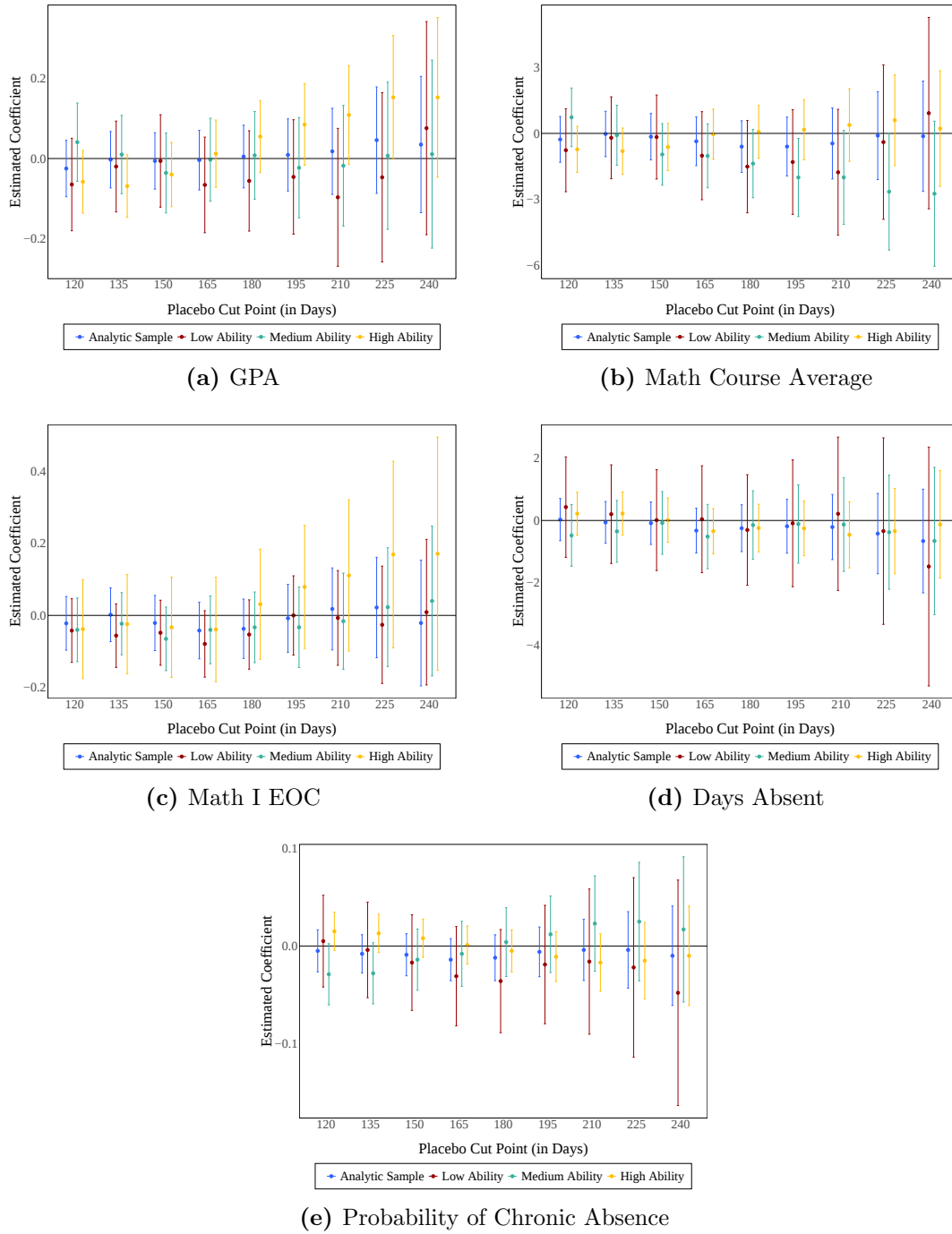
NOTES: The above figure shows discontinuities in the likelihood of 9th grade chronic absence in each listed academic year using a 180-day bandwidth and a quadratic fit.

Figure B5: CDFs of Numeric Grades Over Time



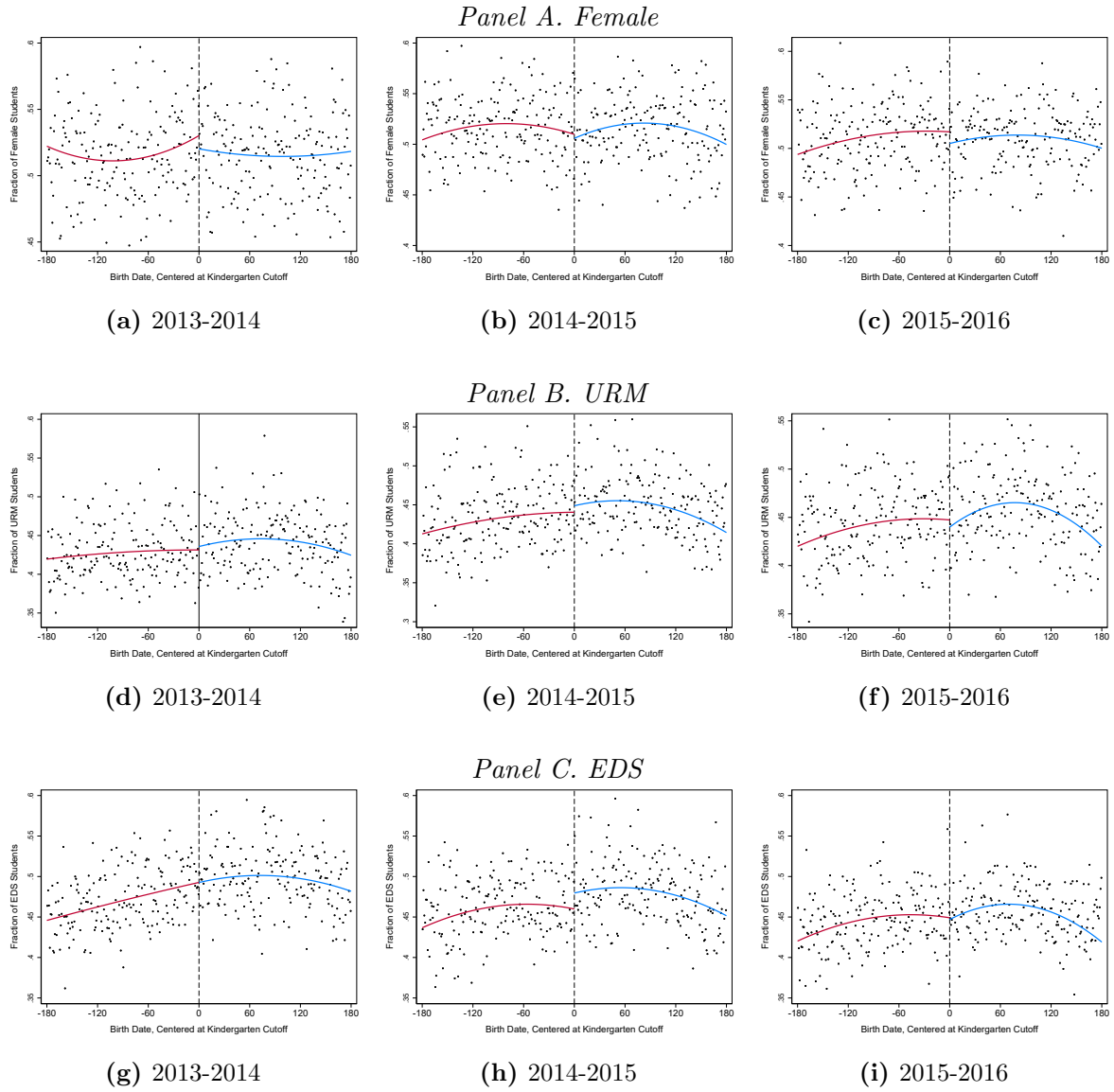
NOTES: The above figure plots the CDFs of numeric grades from 2012-2013 to 2015-16 academic years. CDF plots are presented for core courses and math courses alone in panels (a) and (b). These plots show highlight two features of the data. First, numeric grade distributions are fairly similar across pre-treatment years. Second, in line with [Figure 1](#), the grading policy generated immediate shifts in the numeric grade distributions as shown by the CDF of the 2016 school year.

Figure B6: Placebo Analyses for 9th Grade Outcomes



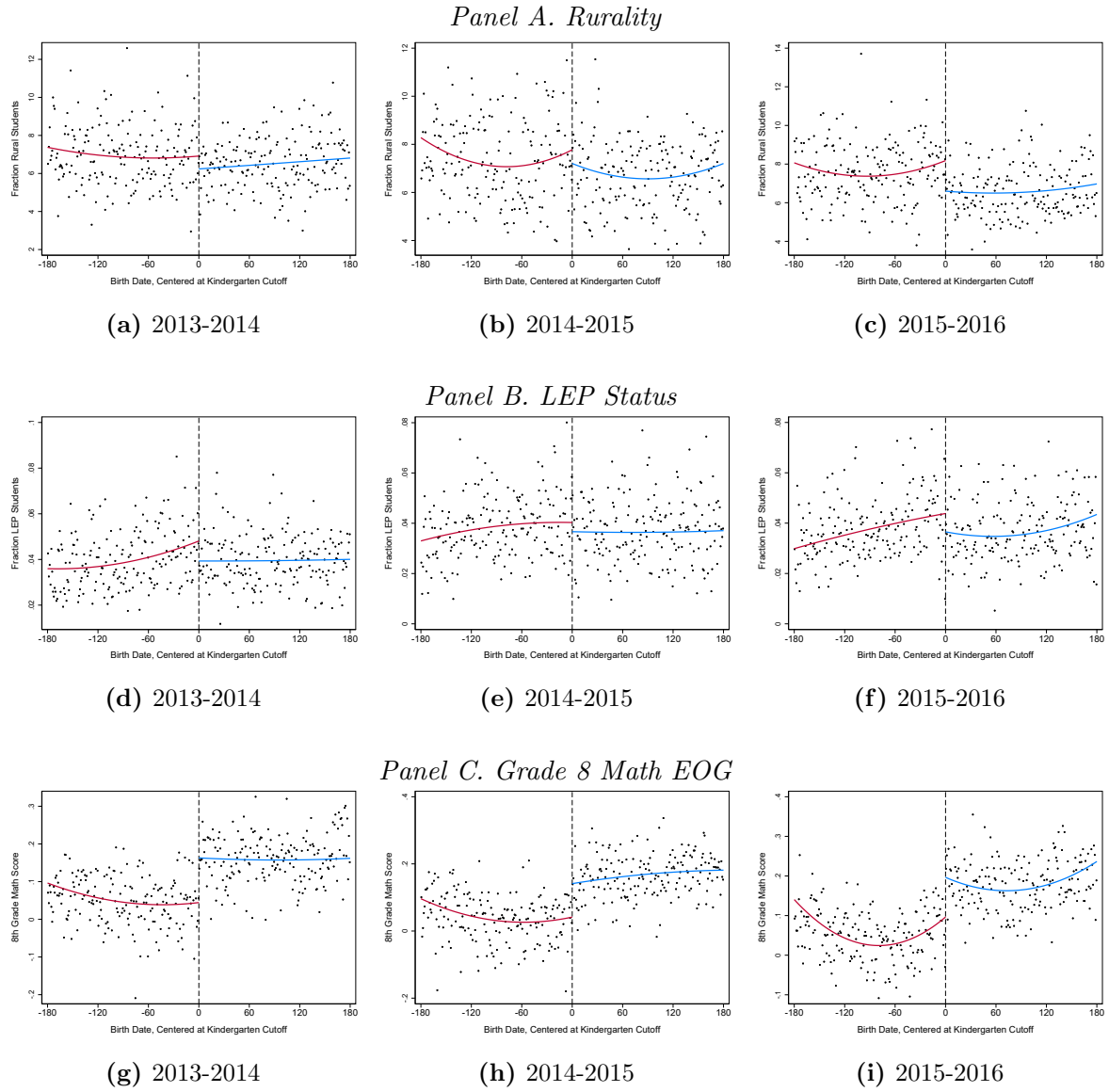
NOTES: The figure above displays sharp difference-in-discontinuity estimates for each of the four main outcomes of interest in 9th grade. Panels (a) through (e) respectively report findings for GPA, 9th grade math course averages, Math I EOC test scores, the number of days absent, and the likelihood to be chronically absent. The placebo cut point (x -axis) refers to the number of days added to the true cut point (day 0 in the main analysis). For each choice of placebo cut point, we show estimates for the main analytic sample and each ability tercile group. We pool academic years 2013-2014 and 2014-2015 to create control windows for each placebo analysis. Robust standard errors at the 95% level are shown.

Figure B7: Regression Discontinuity of Covariates (Part I)



NOTES: The above figure shows discontinuities in share of female, URM and economically disadvantaged (EDS) students (panels A through C) in each listed academic year using a 180-day bandwidth and a quadratic fit.

Figure B8: Regression Discontinuity of Covariates (Part II)

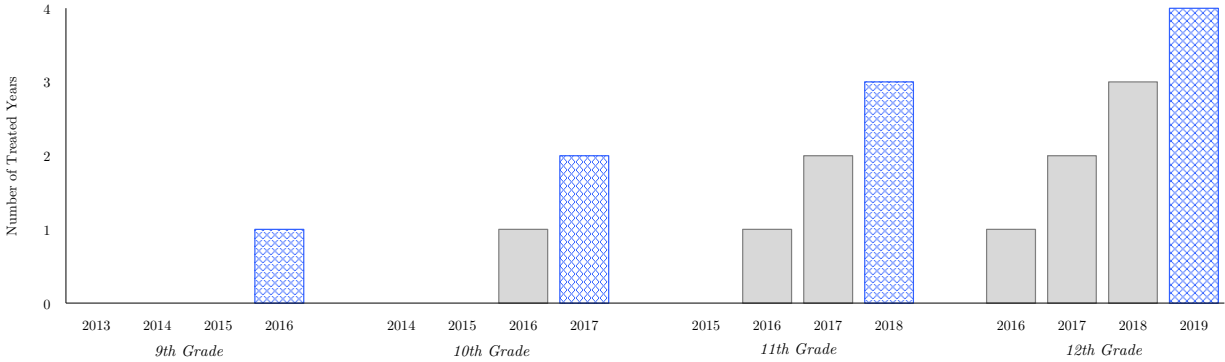


NOTES: The above figure shows discontinuities in rurality, LEP status, and 8th grade Math EOG scores (panels A through C) in each listed academic year using a 180-day bandwidth and a quadratic fit.

C Longer-Run Outcome Contamination

Figure C1 displays the evolution of treatment for students in our analytic sample. For each grade, the first three cohorts correspond to the “control” group and the fourth cohort (blue bar with diamond gradient) corresponds to the “treatment” group. As shown, this design produces clean comparisons in 9th grade, as only 2016 9th graders are exposed to the new policy. However, as our initial 9th grade cohort transitions to 10th grade, one control cohort experiences a year of treatment. If we assume equal cohort sizes, then 10th grade outcomes compare a treated group (100% treatment) to a contaminated control group (17% treatment). By 12th grade, we are comparing a treated group with 100% treatment to a control with 50% treatment.

Figure C1: Dynamic Dosage of Treatment Across High School Years



NOTES: The above figure plots the evolution of treatment “dosage” that occurs as a result of the policy impacting *all* high school students at the time of enactment. The first three years in each group (gray bars) refer to academic years that comprise our control cohorts. The fourth bar in each group (blue bar with diamond gradient) is the treated cohort, based on 9th grade entry. Displayed counts refer to total high school exposure to the 10-point grading scale up through the listed grade.

D Conceptual Framework

We first develop a model of endogenous student responses to grading standards. We construct this simple model to guide the interpretation of our empirical results and provide a stylized prediction for how students may respond differently to changes in grading standards or policies depending on their prior skills and experiences. Unlike prior literature, we specify a model in which students enjoy utility from their *grade point average* rather than their numeric score. Conceptually, a student should not derive additional utility from earning a 95 versus a 94 in a class if both earn that student a grade of A, or 4.0 quality points.¹ Empirically, bunching patterns in the distribution of numeric scores across letter grade cutoffs suggest that these are important markers for students and teachers (see [Figure 1](#)).

D.A Environment

Consider a high school student, i , defined by their latent ability, a_i . We discretize ability according to $a_i \in \{a^\ell, a^h\}$, which refers to low and high ability, respectively. Although we do not formally model the evolution of ability and the impact of socioeconomic input variables ([Todd and Wolpin, 2003](#); [Heckman, 2006](#)), we consider ability as the dynamically-produced realization at the time we observe students in our data. Given this, we assume students know their own type.

Schools are endowed with a grading policy, P , set forth by the district or state. Grading policies map numeric course averages (scores) into quality points. The mean of these quality points forms a student's grade point average (GPA). Formally, $P : \mathcal{S} \times \mathbf{p} \rightarrow \{0, 1, \dots, 4\}$, where $\mathcal{S} \equiv [0, 100] \subset \mathbb{R}$ is the score space and $\mathbf{p} := \{p_A, p_B, p_C, p_D\}$ is the set of cut points. We assume this policy has identical threshold sizes for all grades above an F, meaning $100 - p_A = p_A - p_B$, as well as $p_A - p_B = p_B - p_C$,

¹ In high schools that rank students on the basis of GPA, GPA-dependent rank is one of the most important criteria for college admissions ([Espenshade, Hale and Chung, 2005](#)).

and so on.² As an example, a 10-point policy takes the form

$$P(s_i, \mathbf{p}) = \begin{cases} 4.0, & s_i \in [90, 100] \\ 3.0, & s_i \in [80, 90) \\ 2.0, & s_i \in [70, 80) \\ 1.0, & s_i \in [60, 70) \\ 0.0 & s_i \in [0, 60), \end{cases}$$

where s_i is student i 's numeric final average, or score, in a class. In the above, $p_A = 90$, $p_B = 80$, $p_C = 70$, and $p_D = 60$. In general, a symmetric policy is an n -point one where n denotes the length of each passing grade range.

We assume class scores depend on student ability and exerted effort, e_i . In this paper, we focus on the student's problem of earning a grade in one class, although this model can easily be extended to accommodate a semester's worth of courses.³ Formally, $s_i = s(a_i, e_i)$ for some concave score production function $s(\cdot)$, increasing in a_i and non-decreasing in e_i . We parameterize s_i in the following way:

$$s_i = \mu + \beta a_i + \gamma \ln e_i + \xi_i, \quad \xi_i \sim F.$$

This function satisfies our assumptions for any $(\beta, \gamma) \gg 0$. This form additionally features decreasing marginal returns to effort irrespective of ability. Given our functional assumption on score production, we further impose $e_i \in [1, \bar{e}]$.

The term ξ_i captures shocks to the production of scores. In practice, F can be generalized to any distribution belonging to the class of distributions that have bounded support Ω and are everywhere differentiable along that support. In other words, we assume that $P(\xi_i \in \Omega) = 1$ and that the pdf of F , $f(\cdot)$, exists and is continuous along Ω . We maintain an assumption of boundedness to prevent shocks from taking on extreme values, which would send scores beyond the range $[0, 100]$. In the discussion that follows, we explicitly parameterize the distribution $F \equiv \mathcal{U}(\underline{\xi}, \bar{\xi})$ for $\underline{\xi} := \inf(\Omega)$ and $\bar{\xi} := \sup(\Omega)$.

² This is a simplifying assumption that need not hold for the results to hold.

³ The easiest way to do this would be to assume the semester's utility is the sum of each course's utility. The problem of the student would then change to account for the division of effort across course schedules, rather than the isolated decision to exert effort in any one class.

Students also face costs to exerting effort, c_i , in the form of a convex cost function $c(a_i, e_i)$. In this general setting, effort can refer to, e.g., time spent studying, completing homework, or attending class. We parameterize the cost function according to

$$c(a_i, e_i) = \frac{\kappa e_i}{a_i}.$$

This functional form has the desired properties $\partial c(\cdot)/\partial a_i < 0$ and $\partial c(\cdot)/\partial e_i > 0$ for any $\kappa > 0$. While we proceed under the assumption that cost is linear in effort and production is concave in effort, our analyses would not substantively change if we instead imposed a convex cost function with a linear production function.

Finally, we recognize that the main limitation of this model is that it does not account for the endogenous response of colleges or employers to changes in standards that can also shape student effort decisions.⁴ This is beyond the scope of this model, as it is intended to inform the empirical analysis of the paper that focuses on short-term effects within the K-12 education system.

D.B The Student's Problem

We specify a model in which students enjoy utility from their *grade point average* rather than their numeric score, which departs from the groundwork established in [Betts \(1998\)](#). We parameterize this mapping function $P(\cdot)$ according to:

$$P(s_i) = \sum_{j \in \{A, B, C, D, F\}} \phi_j \mathbf{1}_{\{s_i \geq p_j\}}$$

If, for example, s_i falls in the B range, a student would derive utility $\phi_B + \phi_C + \phi_D + \phi_F$. The ϕ_j term then represents the marginal utility received by earning the next highest letter grade. Without loss of generality, we normalize the return to a grade of F by setting $\phi_F = 0$.⁵ We further impose that $\phi_j > \phi_k$ for any grade $j > k$, meaning that students derive greater marginal utility from accessing higher grades.

Students make effort decisions at the beginning of each semester before ξ_i is realized. As a

⁴ See [Betts \(1998\)](#) and [Chan, Hao and Suen \(2007\)](#) for a deeper exploration of this dimension of response to academic standards.

⁵ For completeness, we note that every grading policy will have $p_F = 0$.

result, they seek to maximize their expected utility from earning a grade,

$$\mathbb{E}[u_i|a_i, e_i] = (\phi_A + \phi_B + \phi_C + \phi_D) \cdot \Pr(s_i \geq p_A) + \dots + \phi_D \cdot \Pr(p_C > s_i \geq p_D) - \frac{\kappa e_i}{a_i}. \quad (\text{D1})$$

Because both high and low ability types incur shocks according to the same distribution, the difference in the levels of ability may generate differences in the *set of feasible grades*, which we denote by $\mathcal{G}^k$ for $k \in \{\ell, h\}$. We assume these sets are distinct, i.e., $\mathcal{G}^\ell \neq \mathcal{G}^h$. In particular, we assume that the environment is defined such that low ability types experience

$$p_C > \underbrace{\mu + \beta a^\ell + \xi}_{\text{lowest possible score}} \geq p_D \quad \text{and} \quad p_A > \underbrace{\mu + \beta a^\ell + \gamma \ln \bar{e} + \bar{\xi}}_{\text{highest possible score}},$$

while high ability types instead experience

$$p_B > \underbrace{\mu + \beta a^h + \xi}_{\text{lowest possible score}} \geq p_C \quad \text{and} \quad \underbrace{\mu + \beta a^h + \gamma \ln \bar{e} + \bar{\xi}}_{\text{highest possible score}} \geq p_A.$$

For either type, the first inequality denotes the infimum of their score set, which occurs when effort is minimized and the lowest production shock is realized; conversely, the second inequality denotes the supremum, obtained whenever effort is maximized and the highest productivity shock occurs. As a result, low ability types have a feasible choice set $\mathcal{G}^\ell = \{B, C, D\}$ and high types instead have $\mathcal{G}^h = \{A, B, C\}$.⁶ We display a graphical representation of these differences in [Figure D1](#).⁷

Following [Equation D1](#), students then solve the following problem:

$$\max_{1 \leq e_i \leq \bar{e}} (\phi_A + \phi_B + \phi_C + \phi_D) \cdot \Pr(s_i \geq p_A) + \dots + \phi_D \cdot \Pr(p_C > s_i \geq p_D) - \frac{\kappa e_i}{a_i}.$$

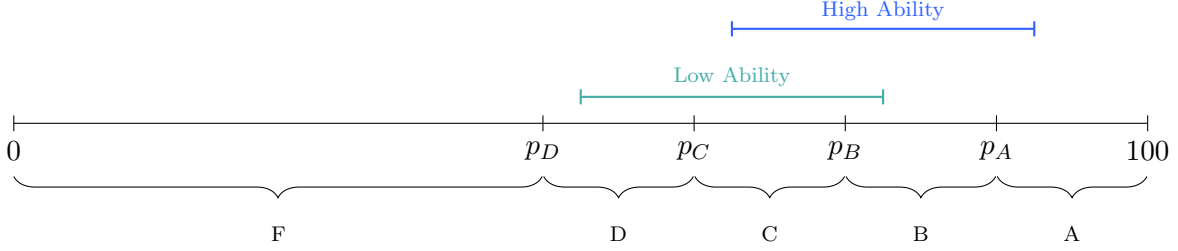
Using the fact that $\xi_i \sim \mathcal{U}_{\{\bar{\xi}, \underline{\xi}\}}$, along with our aforementioned assumptions on feasible grades, it can be shown that low ability types choose an optimal level of effort

$$e_\ell^* = a^\ell \left(\frac{\phi_B + \phi_C}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa},$$

⁶ These feasible sets are somewhat arbitrary. In order to generate our theoretical results, we require only that the set of grades differ. For example, we could consider instead $\mathcal{G}^\ell = \{F, D, C\}$ and maintain our conclusions in this section.

⁷ We recognize that feasible grade ranges can vary across subject domains and across school quality. However, we abstract from these factors to focus on short-term student response to changes in grading standards.

Figure D1: Feasible Grades for Students of Varying Ability Level



NOTES: This figure displays the difference in feasible scores for students of high and low ability type. The corresponding grade range is captured by the underbraces. For example, students earning a score between p_A and 100 earn a grade of A. The range for both types has identical length. The difference in the respective beginning or end point between the two types has a value of $\beta(a^h - a^\ell)$. This figure further demonstrates an example in which $\mathcal{G}^\ell = \{B, C, D\}$ and $\mathcal{G}^h = \{A, B, C\}$.

while high ability students instead choose

$$e_h^* = a^h \left(\frac{\phi_A + \phi_B}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa}.$$

From the above, it follows that $e_h^* > e_\ell^*$, which brings us to our first result:

Result 1: For any given grading policy P and collection of students $I \equiv \bigcup_i$ defined by their distinct ability levels a_i , a pair $\{j, k\}$ that satisfies $a_j > a_k$ and either $\mathcal{G}^j \not\equiv \mathcal{G}^k$ or $\mathcal{G}^j \equiv \mathcal{G}^k$ will also imply $e_j^* > e_k^*$; that is, students with higher ability levels optimally choose to exert a higher level of effort compared to lower ability students regardless of whether the difference in ability generates a difference in the set of feasible grades.

D.C The Effects of a Policy Change

We now consider the effects of a state or district changing their grading policy P in favor of a new policy P' with corresponding cut points $\{p'_A, \dots, p'_F\}$. For tractability of our empirical setting, we assume in this discussion that policy P' is more lenient than P , meaning the value of each cut point is lower.⁸ Suppose P' shifts p_A by $d > 0$. In other words, the district ends their use of an n -point grading scale in favor of an $(n + d)$ -point grading scale, yielding the following relationships between new cut points and old ones:

$$p'_A = p_A - d, \quad p'_B = p_B - 2d, \quad p'_C = p_C - 3d, \quad p'_D = p_D - 4d.$$

⁸ The opposite results will hold if instead P' is less lenient.

Perhaps unintentionally, this design generates larger changes for lower grades. This means that, e.g., students that strive for C's experience a relatively larger relaxation in standards than students that strive for A's.

We finally assume that $\xi_i \perp P$, which implies that $s_i \perp P$. This is equivalent to an assumption that scores are produced exogenously to grading policies, which allows us to directly compare the production of scores between policies. One possible violation to this would be if teachers differentially curved grades in response to policy changes.⁹ We do not include the dimension of teacher response in our model because our research design is able to overcome this identification challenge.

Based on optimal effort decisions, we show the effects of a policy change are both (1) ambiguous for a given student and (2) potentially heterogeneous between types of students. To demonstrate this, we first solve the problem for low ability types in isolation and then introduce results including high ability types. Under the initial n -point policy P , low types experienced $\mathcal{G}^\ell \equiv \{B, C, D\}$. The policymaker's selection of d that forms the new $(n + d)$ -point policy P' generates the new feasible grade set $\mathcal{G}^{\ell'}$. This new feasible set can be identical to $\mathcal{G}^\ell$ and generate no change in student effort, which we call a *stationary* policy. Alternatively, the policy can eliminate the lowest possible grade while maintaining the highest possible grade (i.e., $\mathcal{G}^{\ell'} = \{B, C\}$), which leads student to decrease their effort. We term this a *contractionary* policy. Finally, the policy can introduce a new highest possible grade (i.e., $\mathcal{G}^{\ell'} = \{A, B, C\}$ or $\mathcal{G}^{\ell'} = \{A, B, C, D\}$), leading students to increase their effort. We call this an *expansionary* policy.

Under a stationary policy, it will necessarily be the case that $\Delta e_\ell^* = 0$. However, whenever the lenient policy is contractionary, students will reduce their effort absolutely:

$$e_\ell^{*'} = a^\ell \left(\frac{\phi_B}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa} < a^\ell \left(\frac{\phi_B + \phi_C}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa} = e_\ell^* \quad \implies \quad \Delta e_\ell^* < 0.$$

Conversely, an expansionary policy increases effort, regardless of whether the lower bound is changed.¹⁰ Formally, for the case in which $\mathcal{G}^{\ell'} = \{A, B, C\}$,

$$e_\ell^{*'} = a^\ell \left(\frac{\phi_A + \phi_B}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa} > a^\ell \left(\frac{\phi_B + \phi_C}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa} = e_\ell^* \quad \implies \quad \Delta e_\ell^* > 0.$$

⁹ While we do not model the role of teachers' discretion in assigning grades, further work could expand our model to incorporate this in the style of [Diamond and Persson \(2016\)](#).

¹⁰ In the most extreme case, which we ignore as unrealistic, a policy could make the set of feasible grades singular, which would actually decrease effort to its minimum.

Therefore, a student whose feasible grade set does not include either an A or an F has an entirely ambiguous response to a lenient grading policy. The effect will depend on both the size of the policy d and the relative span of their feasible score set.

In the case of the high ability student, the effects are less ambiguous. Due to the fact that $\mathcal{G}^h$ includes an A, no lenient policy P' can have an expansionary effect on high ability students. Following the previous analysis, if $\mathcal{G}^{h'} = \{A, B\}$, then $\Delta e_h^* < 0$; otherwise, high ability students will not change their effort in response to P' . This leads us to our next result:

Result 2: A new grading policy P' which is more lenient than the currently enacted policy P will have a non-positive effect on the effort exerted by the highest ability students. The effect for lower ability students is ambiguous and depends on both the magnitude of the policy's leniency and the relative capability of these students.

We conclude this section by graphically demonstrating the ambiguity of the effects of these types of policies, as depicted in [Figure D2](#). In each of the six panels, we consider the same selection of d but vary the relative location of the score set for each ability type.¹¹

Note that, by construction, P' induces a higher expected grade for all students. If effort embodies different measures of student engagement like attendance or study hours, then effort is a productive input in the accumulation of human capital. A reduction in effort that coincides with an increase in student GPA would suggest an artificially-driven inflation of grades. However, lowering standards can result in a net increase in effort across students, as evident in [Figure D2a](#). In this instance, high-achieving lower ability students are able to earn previously unobtainable grades. At the same time, d is chosen to be small enough so that high ability students do not reduce their effort. Importantly, this policy represents a possible way to mitigate achievement gaps.¹²

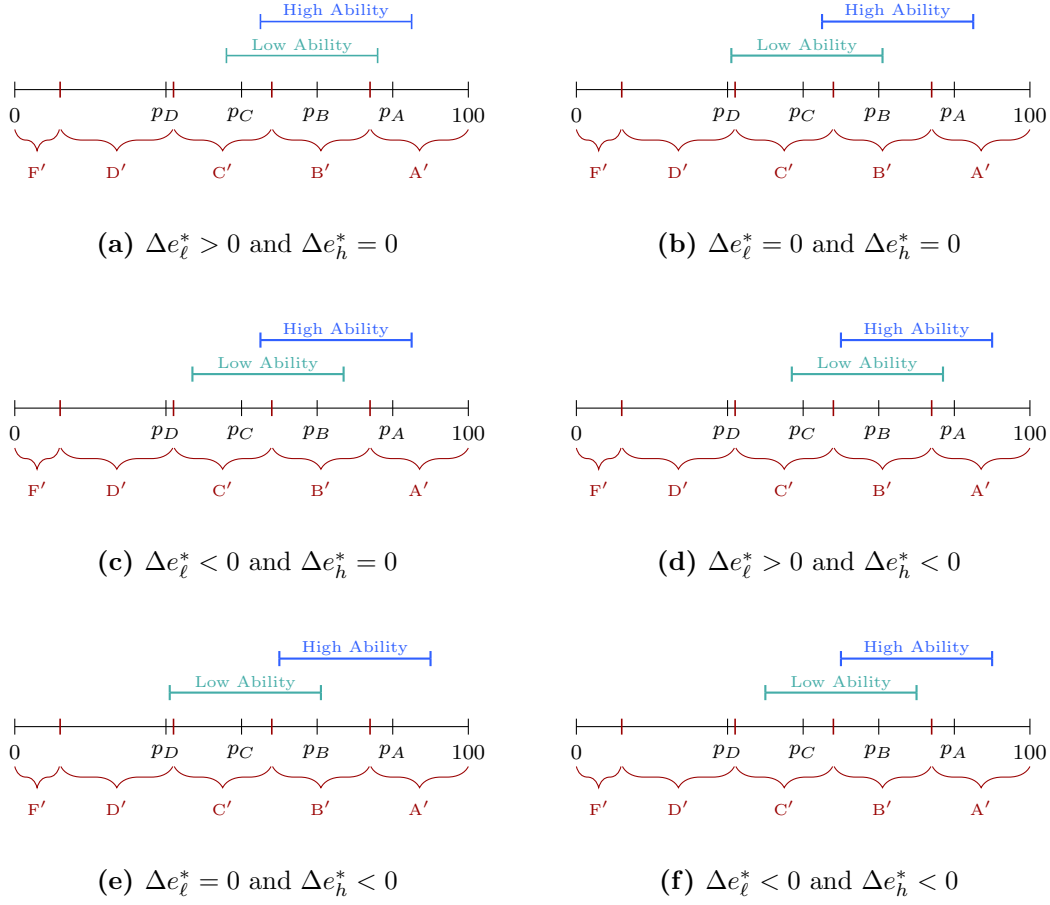
Alternatively, lowering grading standards could exacerbate the achievement gap. For example, the policy enforced in [Figure D2c](#) results in a widening of the achievement gap (low ability students reduce their effort while high ability students maintain their effort) despite the fact that an identical policy reduced the gap in [Figure D2a](#).

Overall, these disparate predictions point to the importance of policy design that is relevant to the target student population. These predictions also help to explain why the literature on grade inflation has mixed—and at times incongruous—findings. The key takeaway of this model is that

¹¹ We can show the same six outcomes if we instead fix the distribution of students but allow the choice of d to vary.

¹² [Figure D2e](#) also depicts a possible way to reduce the achievement gap, albeit without improving the human capital accumulation of any group of students. Similarly, [Figure D2d](#) showcases a policy which reduces engagement among high ability students while boosting effort among low ability students.

Figure D2: Ambiguity in Policy Effects for Different Ability Types



NOTES: The above figures illustrate different policies and ability distributions which could result in heterogeneous responses among students. In each panel, the original cut points for policy P are denoted in black. The new cut points corresponding to policy P' are denoted by the red lines, with the corresponding new grade regions outlined by the red braces and prime letters.

student response depends in part on the *magnitude of academic leniency* induced by the policy and the *discrepancy and spread* of the student score distribution. The latter is a function of students' abilities and the return to their effort. As such, a relaxation of grading standards may lead students to decrease their investments in school in some contexts (Betts and Grogger, 2003; Figlio and Lucas, 2004; Babcock, 2010; Nordin, Heckley and Gerdtham, 2019; Hvidman and Sievertsen, 2021), while at the same time motivate and benefit students in other contexts (Dee et al., 2016; Ahn et al., 2024; Minaya, 2020). As educators and policymakers seek to change grading standards in their school districts, it is important that they understand that the heterogeneity and the direction of the effect will depend on both the score distribution of their student population and the magnitude of their grade change.