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 Słoczyński, 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)

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

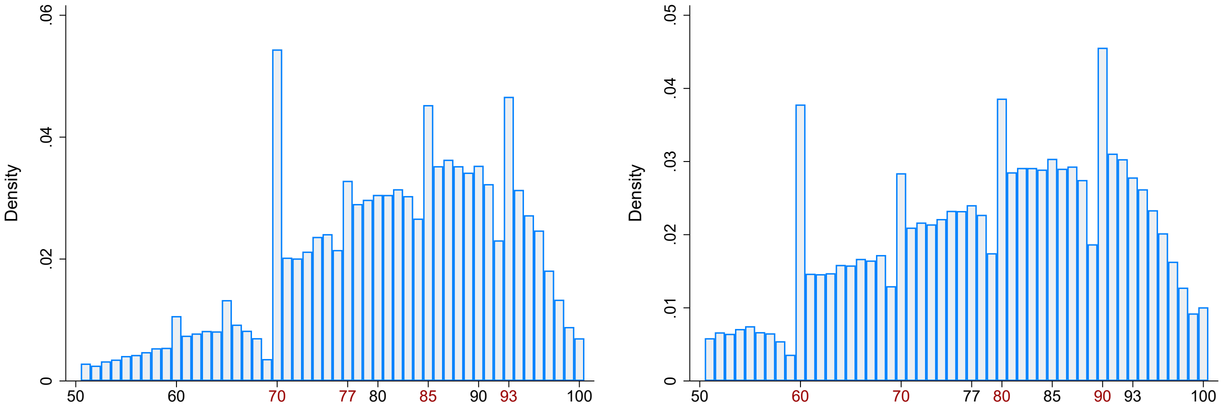
Table A3: Alternative Specifications (Part II)

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

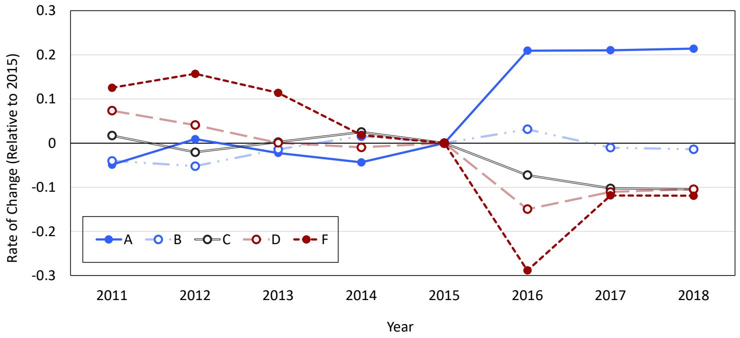


(a) 2014

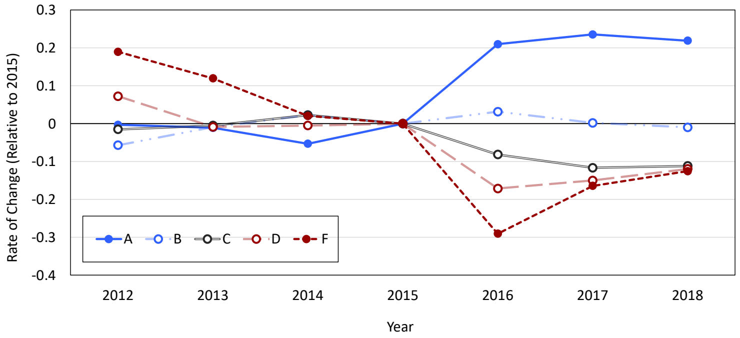
(b) 2017

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.

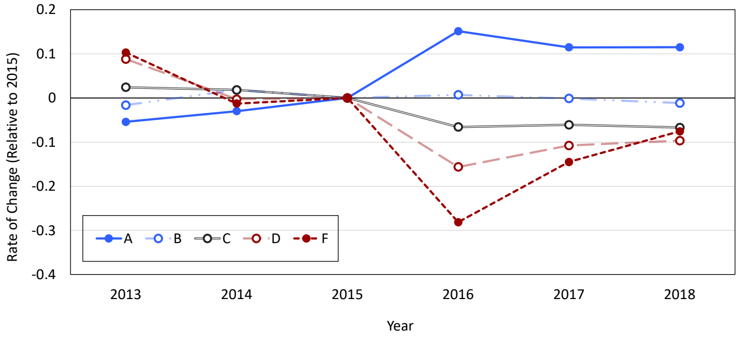
Figure B2: Change in Older Grade Letter Grade Shares, by Year



(a) 10th Grade



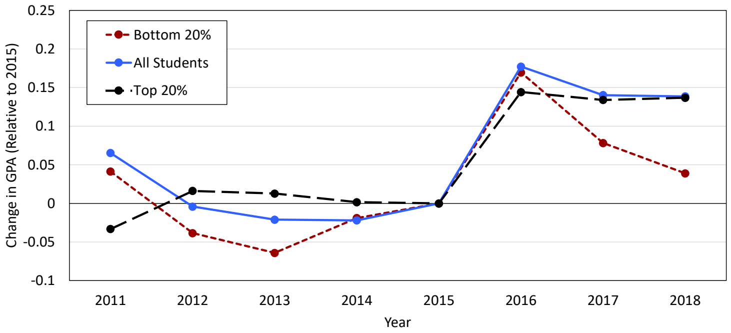
(b) 11th Grade



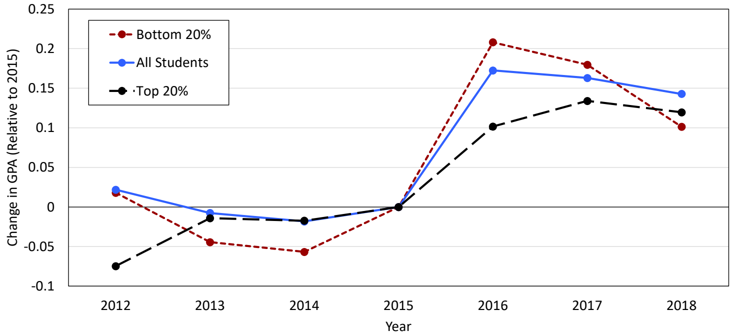
(c) 12th Grade

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).

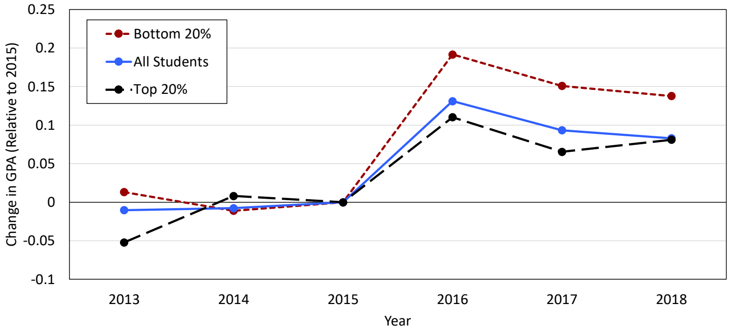
Figure B3: Change in Older Grade GPAs, by Year



(a) 10th Grade



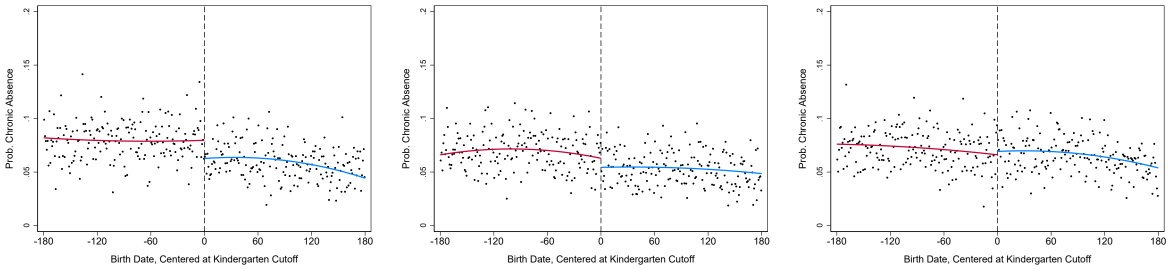
(b) 11th Grade



(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).

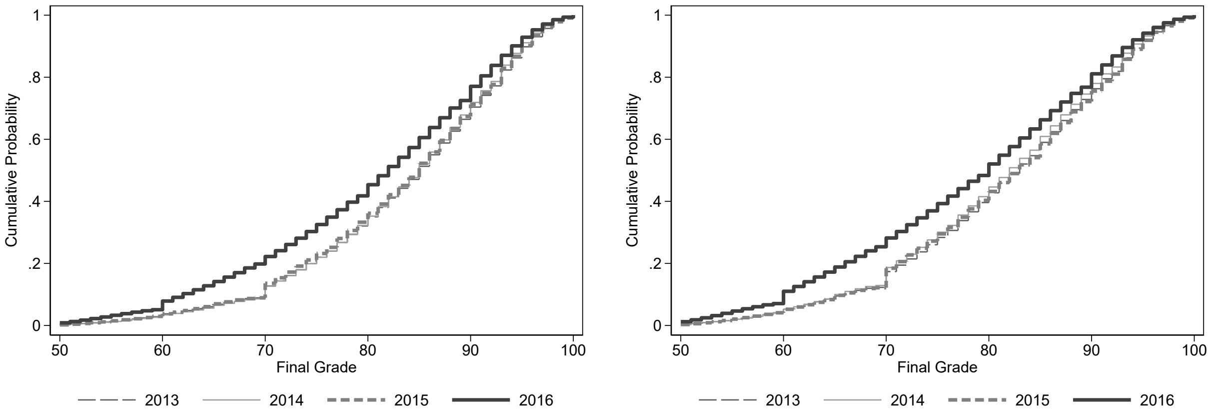
Figure B4: Regression Discontinuity of Chronic Absences Across Academic Years



(a) 2014 (Control) **(b)** 2015 (Control) **(c)** 2016 (Treatment)

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.

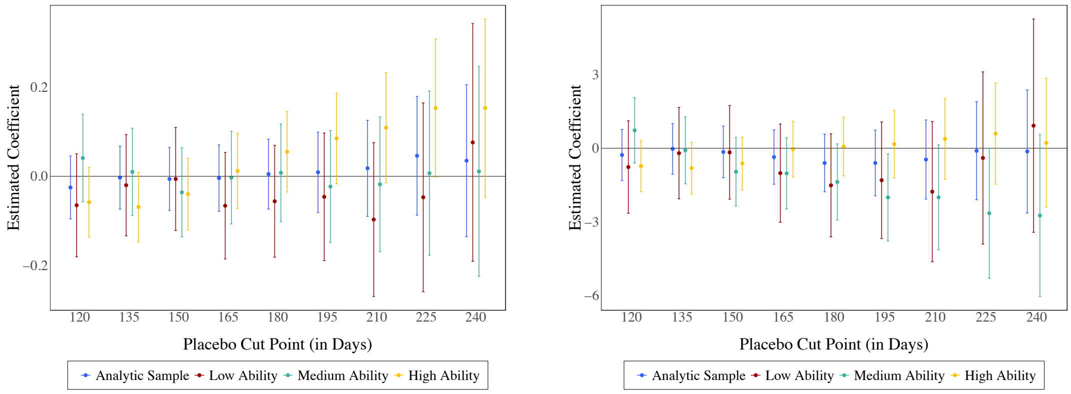
Figure B5: CDFs of Numeric Grades Over Time



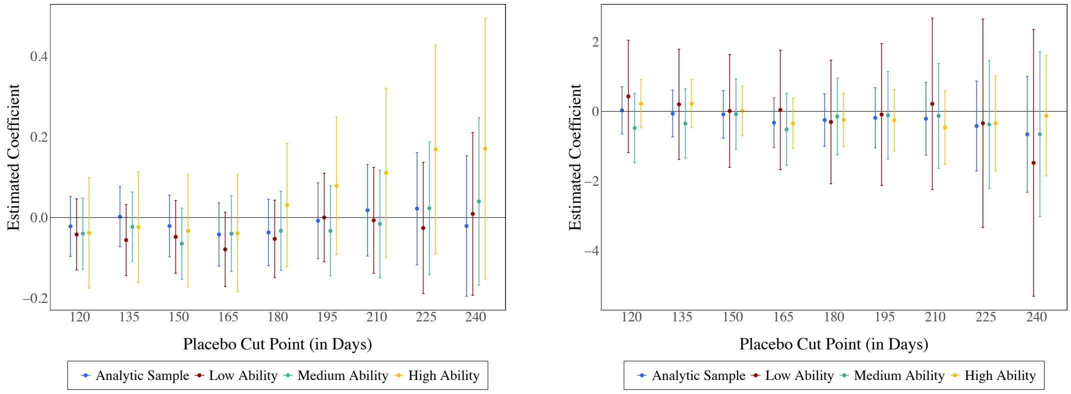
(a) Core Courses **(b)** Math

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.

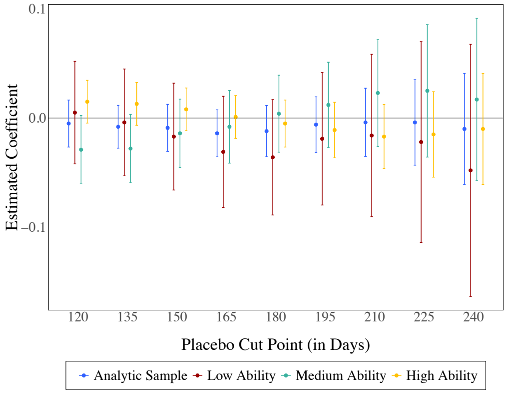
Figure B6: Placebo Analyses for 9th Grade Outcomes



(a) GPA **(b)** Math Course Average



(c) Math I EOC **(d)** Days Absent

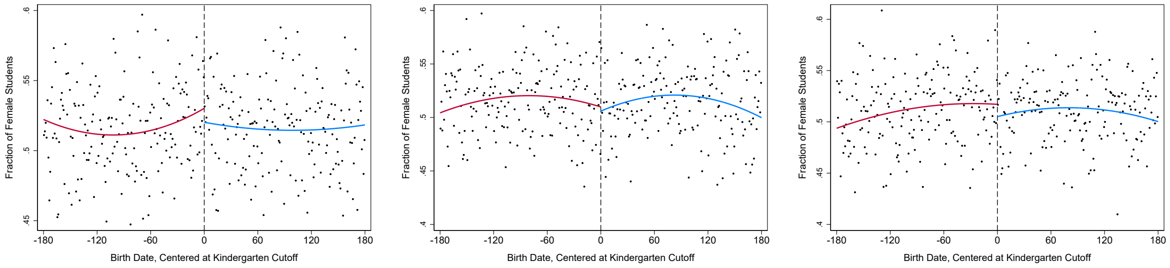


(e) Probability of Chronic Absence

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.

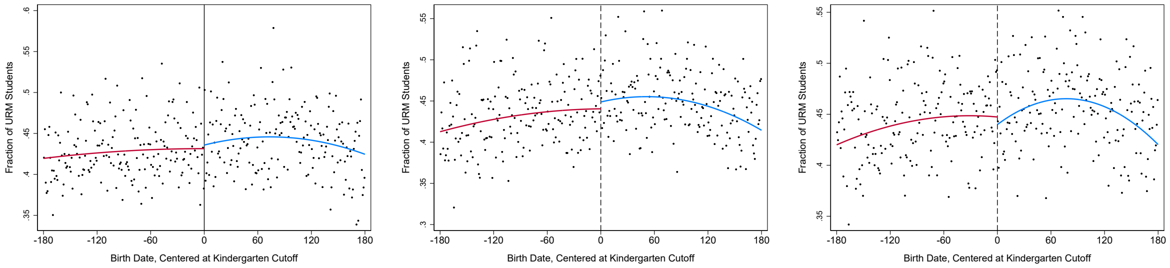
Figure B7: Regression Discontinuity of Covariates (Part I)

Panel A. Female



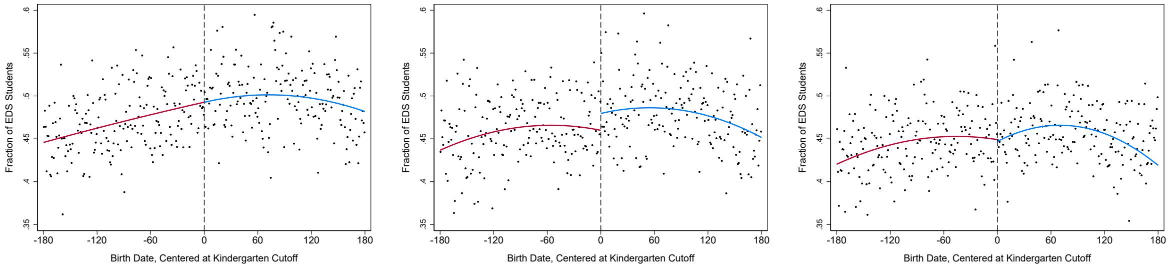
(a) 2013-2014 **(b)** 2014-2015 **(c)** 2015-2016

Panel B. URM



(d) 2013-2014 **(e)** 2014-2015 **(f)** 2015-2016

Panel C. EDS

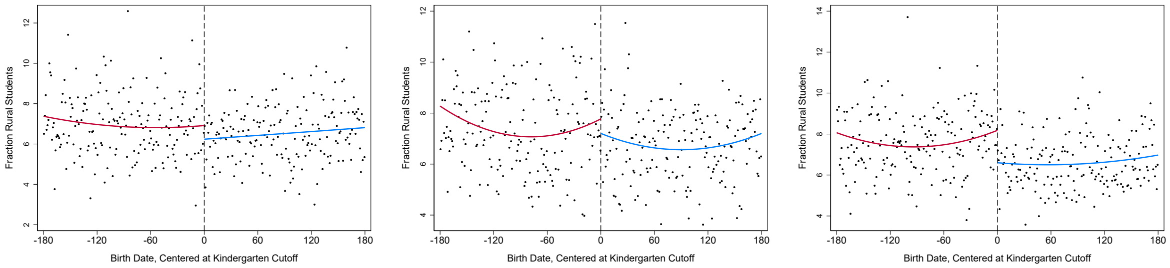


(g) 2013-2014 **(h)** 2014-2015 **(i)** 2015-2016

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)

Panel A. Rurality

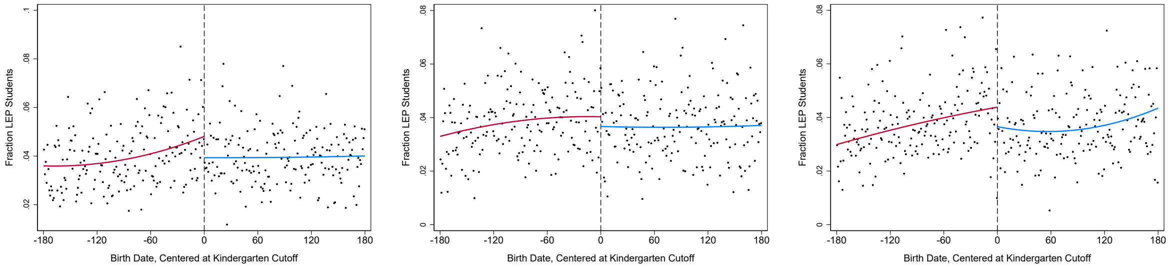


(a) 2013-2014

(b) 2014-2015

(c) 2015-2016

Panel B. LEP Status

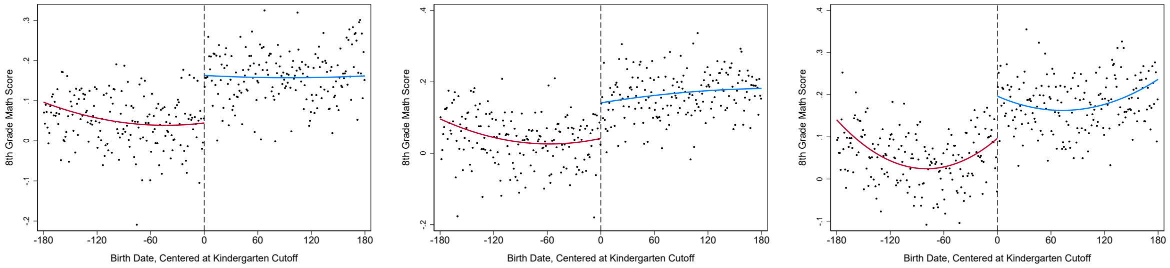


(d) 2013-2014

(e) 2014-2015

(f) 2015-2016

Panel C. Grade 8 Math EOG



(g) 2013-2014

(h) 2014-2015

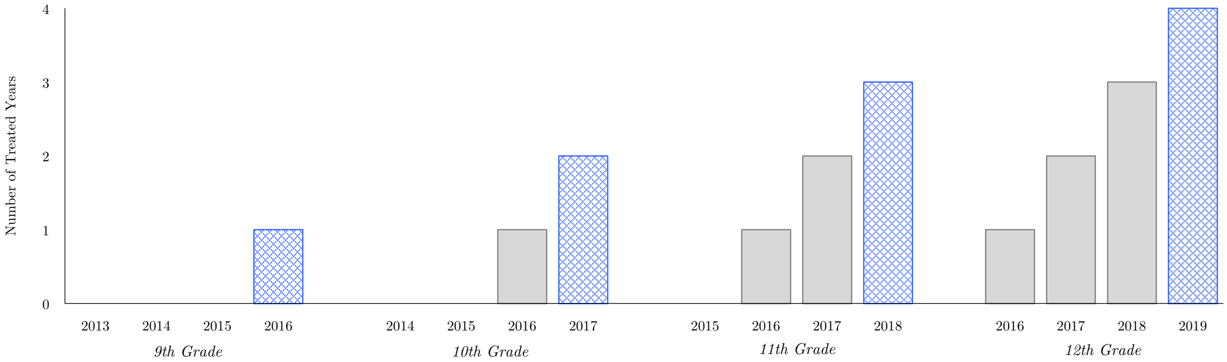
(i) 2015-2016

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 patters 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, $P : \mathcal{S} \times \boldsymbol{p} \to \{0, 1, ..., 4\}$, where $\mathcal{S} \equiv [0, 100] \subset \mathbb{R}$ is the score space and $\boldsymbol{p} := \{p_A, p_B, p_C, p_D\}$ 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

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

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 $(\beta, \gamma) >> 0$. 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:

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

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,

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

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., $\mathcal{G}^\ell \not\equiv \mathcal{G}^h$. In particular, we assume that the environment is defined such that low ability types experience

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

while high ability types instead experience

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

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:

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

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

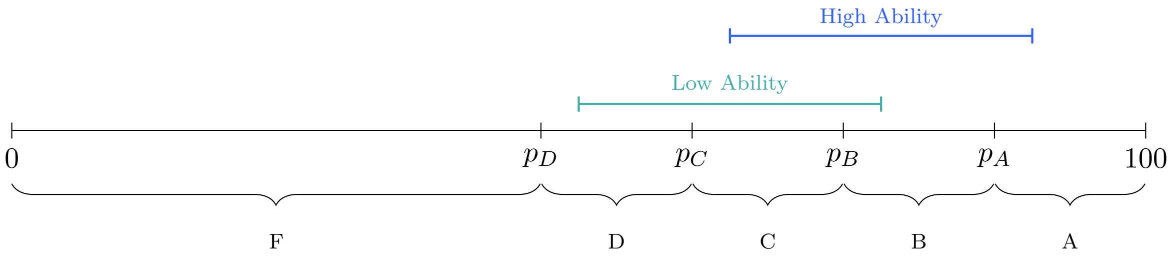
$$e^*_\ell = a^\ell \left(\frac{\phi_B + \phi_C}{\bar{\xi} - \underline{\xi}} \right) \frac{\gamma}{\kappa},$$

---

[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 $G^\ell = \{F, D, C\}$ 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 $\{p'_A, ..., p'_F\}$. 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:

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

---

[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:

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

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

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

---

[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.

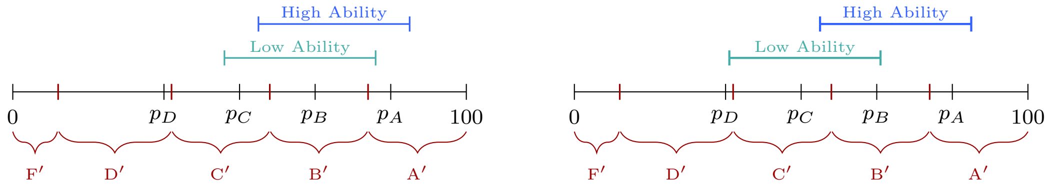
Overall, these disparate predictions point to the importance of policy design that is relevant to the target student population. These predictions also help to explain why the literature on grade inflation has mixed—and at times incongruous—findings. The key takeaway of this model is that

---

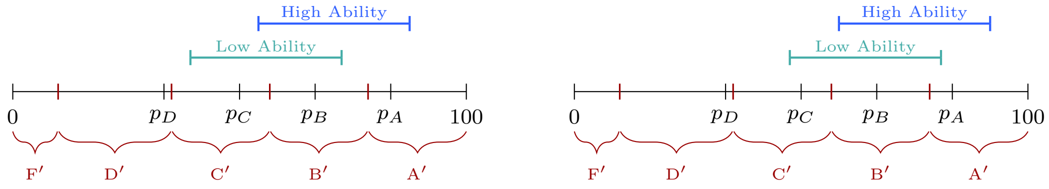
[11] We can show the same six outcomes if we instead fix the distribution of students but allow the choice of d to vary.

[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.

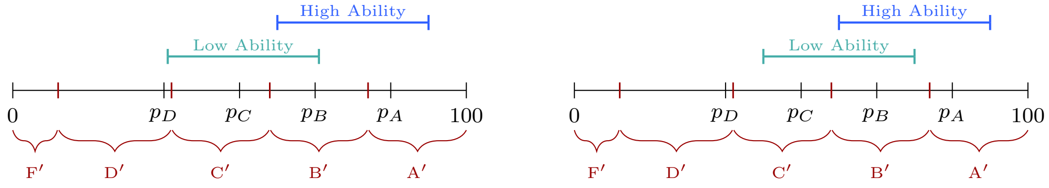
Figure D2: Ambiguity in Policy Effects for Different Ability Types



(a) $\Delta e_\ell^* > 0$ and $\Delta e_h^* = 0$ **(b)** $\Delta e_\ell^* = 0$ and $\Delta e_h^* = 0$



(c) $\Delta e_\ell^* < 0$ and $\Delta e_h^* = 0$ **(d)** $\Delta e_\ell^* > 0$ and $\Delta e_h^* < 0$



(e) $\Delta e_\ell^* = 0$ and $\Delta e_h^* < 0$ **(f)** $\Delta e_\ell^* < 0$ and $\Delta e_h^* < 0$

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.