From Persistence to Performance: The Effects of Elite High Schools on Retaking and Testing

Marco A Acosta*

September 9, 2025

(Latest Version)

This is a preliminary version. Please do not cite or circulate without author's permission.

Abstract

This paper investigates the effects of elite school eligibility on retake probability and effort in the context of the centralized admission system in Mexico's metropolitan area, where upper secondary schools predominantly use entrance exam test scores to determine student admissions. Notably, elite high schools require an additional criterion: a middle school GPA of at least 7 out of 10. Using this policy rule, I employ regression discontinuity estimates comparing students that are eligible and ineligible to attend elite schools. I find that eligible students are more likely to retake the entrance exam and perform better. In addition, their families increase spending on private preparatory courses, and they do not show significant differences, with respect to ineligible students, in self-reported anxiety, aggression, depression, or attention spam indicators.

Keywords: effort, regression discontinuity design, student achievement, retaking.
JEL classification: I21, I25, I28.

*Email address: macostac@iu.edu.

I am especially grateful to Michael Kaganovich and Ruli Xiao for their guidance and mentorship throughout this project. I also thank Jack Berger, Jarrod Burgh, Andrew Dustan, Vicente Garcia, Rafa de Hoyos, Maria Padilla, and Miguel Urquiola for valuable feedback and suggestions, and *Xaber* for providing access to the data used in this study. All remaining errors are my own.

1 Introduction

To reduce the extent of enrollment manipulation and provide fair and efficient assignments, educational authorities around the world have established centralized school choice mechanisms in the form of report-specific priority + cutoff, where priority is established primarily by an entrance exam test score (Agarwal and Somaini, 2020).[1] The change from decentralized to centralized admission prompted students and families to devote substantial time and financial resources to test preparation in order to gain access to prestigious schools and colleges, even though many ultimately do not secure a place (Dang and Rogers, 2008), while recent evidence shows that during the study spam associated to centralized exams students face adverse mental health effects (Yatkin et al., 2023; Yang et al., 2023). To mitigate these pressures, some governments have acted to curb excessive study hours, to safeguard students well being, and lower the financial burden on families, a step also intended to level the playing field for gaining access to elite schools (Qian et al., 2024; Bjork, 2011). Understanding whether students' exam preparation effort is driven by aspiring to enroll into elite schools is essential for guiding policies, especially in light of the evidence that admission test preparation does not add lasting educational value.[2]

This study provides evidence on the extent to which the prospect of attending elite schools shapes the incentives of students to exert effort in the entrance exam of the centralized admission process to upper secondary school in the Metropolitan Area of Mexico. The main challenge in establishing this relationship lies in isolating the effect this incentive from other factors that may also influence test scores. I overcome this challenge for three reasons. First, the seat allocation mechanism is implemented through a serial dictatorship algorithm that constructs priority indexes solely based on students' entrance exam test scores. This ensures that students can only improve their chances of admission to a more preferred school by raising their test performance. Second, an administrative rule prevents students with middle school GPAs below seven from accessing the elite school market, effectively dividing applicants into two distinct groups: those eligible for elite allocations and those barred from such

[1]Some countries that use a priority+cutoff centralized matching mechanism in *high school* include: China (Wang and Zhou, 2020), France (Hiller and Tercieux, 2014), Ghana (Ajayi, 2024), Mexico (Dustan et al., 2017), Romania (Pop-Eleches and Urquiola, 2013), Singapore (Teo et al., 2001), Turkey (Akyol and Krishna, 2017), and the United States (Abdulkadiroğlu et al., 2014).

For *university* admissions, examples include: Australia (Artemov et al., 2017), Brazil (Machado and Szerman, 2016), Chile (Rios et al., 2021), China (Chen and Kesten, 2017), Hungary (Biró, 2007), Norway (Kirkeboen et al., 2016), Spain (Luflade, 2018), and Turkey (Karadal et al., 2013).

[2]Admission systems in other countries, that do not rely on centralized matching, may also place a heavy weight on standardized entrance exams. For example, in the United States, the SAT and ACT historically played a central role in college admissions, while in South Korea, the College Scholastic Ability Test (CSAT) is the primary determinant of university admission. This paper also speaks to such systems.

opportunities. Third, the elite segment of the market is clearly identifiable, as it consists of high schools affiliated with the most prestigious universities in the country, namely the Instituto Politécnico Nacional (IPN) and the Universidad Nacional Autónoma de México (UNAM). Together, these institutional features provide a credible setting to estimate the causal impact of elite school eligibility on student effort using regression discontinuity (RD) design, and also enable an examination of related outcomes such as families' financial investment in preparatory courses and socioemotional responses, including, anxiety, aggression, depression and attention spam.

The analysis centers on retakers, as students undertaking the high school entrance examination for the first time remain uninformed of their final middle school GPA due to the enrollment process's timeline. This is evidenced by the fact that the majority of first-time test-takers below the eligibility threshold continue to include elite schools, despite not being eligible to attend due to their GPA status. Conversely, candidates retaking the examination, having full awareness of the GPA criteria and positioned below the seven threshold, do not include elite schools within their options.

Using retakers subpopulation poses a significant identification threat, as students self-select into retaking the entrance exam, potentially biasing results. This occurs because students with initial aspirations to attend elite schools, yet hindered by low GPAs, may opt out of retaking, while those above the threshold might retry, thus generating unbalanced comparison groups at the middle school cutoff. To address this issue, I implement different econometric approaches that differ in their identification assumptions and target estimands. Within the RD design framework, I employ three approaches to estimate the Local Average Treatment Effect (LATE). The first is the canonical RD desing (Hahn et al., 2001). The second explicitly models selection to estimate both the extensive margin, the change in the retake probability induce by the policy, and the intensive margin, the change in test scores among retakers (Dong, 2019). The third uses the first difference in the outcome variable to restore smoothness, measuring test scores gains, at the middle school cutoff (Lee and Lemieux, 2010). Beyond RD design, I employ strategies that rely on unconfoundedness, which allow for the estimation of conditional average treatment effects (CATE), either by controlling for observable characteristics using linear random forest (Athey et al., 2019) or by accounting for unobservable characteristics, proxied by the proportion of elite schools in students' high school applications (Dale and Krueger, 2014).

The analysis draws on rich administrative and survey data from the COMIPEMS (Comisión Metropolitana de Instituciones Públicas de Educación Media Superior, by its acronym in Spanish) centralized admission system in 2015 and 2016, which regulates access to public high schools in the metropolitan area of Mexico. Each year, approximately 300,000 students participate in the entrance exam, commonly referred to as COMIPEMS exam, ranking up 20 of their preferred schools while admissions are determined in sequential form by test scores subject to school capacity constraints. The data combine detailed information from the admission process, including test scores, and individual level survey responses covering demographic characteristics, socioeconomic background, and students behavior.

The results consistently show that the GPA policy increased student effort, as reflected in higher test scores among those just above the eligibility threshold. Canonical RD design estimates increases in test scores of 0.23 standard deviation (sd), while the gains specification that account for prior performance yield smaller but still significant results of 0.10 sd. The decomposition into extensive and intensive margins indicates that the policy increased the likelihood of retaking the exam by 2.6 percentage points, with improvement in test scores among retakers of the same magnitude as the canonical RD design. Sharp bounds for the intensive margin range from close to zero to as high as 0.37 sd. Estimate of observables confirms positive effects of around 0.10 sd, while controlling for unobserved factors through student application choices estimates the impact at 0.18 sd. To provide a sense of magnitude, translating these estimates into a *business as usual* school year (Evans and Yuan, 2019), an effect of 0.10 sd corresponds to six months of additional schooling, or two fold the estimated impact of the full-time school program, that extended the school day by four hours in elementary education, on the high school entrance exam (Cabrera-Hernández et al., 2023).[3]

With respect to family investment in preparatory courses, the results indicate that families whose children are eligible to attend elite schools devote more money in additional preparation. Eligibility increases enrollment in private preparatory courses by 15 percentage points. In contrast, ineligible students are about 14 percentage points more likely to participate in preparatory courses offered directly at their middle schools, which are free of charge and thus provide a costless alternative to the paid private options. The observed pattern indicates that although both groups engage in preparatory efforts, one group allocates a greater degree of financial resources, while the other depends on complementary school-based resources. Private tutoring participation remains similar across eligibility groups.

Regarding socioemotional indicators. The results show that there is no statistically significant impact of the middle school policy in self-reported anxiety, aggression, depression, or attention problems between eligibility groups. However, this result should be interpreted

[3]The test scores increases in the high school entrance exam stem from targeted exam preparation in response to the GPA policy, rather than sustained instructional exposure.

with caution. The null effects may reflect the fact that the analysis focuses on students near the lower end of the achievement distribution or that the survey instrument was not specifically designed to capture mental health outcomes. With more specialized questionnaires, or among higher achieving students, it is possible that the results would differ, as documented in other countries where eligibility for elite schools has been associated with greater stress and poorer mental health (Chen et al., 2022; Xu and Lee, 2023).

This paper contributes to the general economic literature on incentives and effort (Gneezy and Rustichini, 2000; DellaVigna and Pope, 2018; Ehrenberg and Bognanno, 1990), and more specifically to the identification of effort in education. Despite the central role that effort plays in shaping academic outcomes, much of the literature abstracts from modeling it explicitly, often relegating effort to unobserved heterogeneity, since effort is rarely captured in administrative data (e.g., study hours, practice examinations, and similar activities) and is difficult to disentangle from ability or other unmeasured inputs. Some exceptions are Stinebrickner and Stinebrickner (2008) that measures effort relying on a rich data set and using an instrumental variable approach and evidence from Gneezy et al. (2019) that shows that the gap between U.S. students and top Chinese performers narrows substantially when both groups are offered identical monetary incentives on the PISA test. Another tangible contribution of this paper is to demonstrate that the choice set students face directly shapes exam effort. This finding provides support in favor of studies that rely on student application choices to control for unobserved heterogeneity (Dale and Krueger, 2014; Abdulkadiroğlu et al., 2020).

A second contribution is to the literature on family investment in education to attend prestigious schools (Liu and Neilson, 2011; Zhang, 2013; Dang and Rogers, 2008; Entrich, 2017; Kim and Lee, 2010), by establishing a causal link between the desirability of elite schools and the investments families make in preparatory courses and related educational expenditures, whereas most existing studies provide more anecdotal or correlational evidence.

A third contribution documents higher retake rates among students eligible for elite schools relative to those ineligible. This finding aligns with evidence on retaking to improve test scores in pursuit of more selective placements (Goodman et al., 2020; Krishna et al., 2018).

A limitation of this study is that the estimates identify LATE rather than the average treatment effect (ATE). Thus capturing the effort for students at the margin of eligibility and may not generalize to the broader population of applicants. Nonetheless, it is reassuring that the results remain statistically significant even when identified at the lower end of the GPA distribution. Moreover, while the institutional setting is specific to Mexico City, the underlying mechanism is relevant to many other countries where high-stakes entrance exams determine admission to selective schools and universities. In such contexts, the magnitude of effort responses is likely to be even more pronounced, as students and their families face intense pressure to secure access to elite institutions.

The structure of the paper is organized as follows. Section 2 provides a description of the institutional context of the Mexican high school educational system and admission process. Section 3 discusses the data and presents the methodology. Section 4 shows the estimation results. Section 5 explores the validity of these results under the RD design framework. Section 6 examines the impact of the policy in financial investment and socioemotional indicators, and Section 7 concludes.

2 Institutional Context

In the transition from middle school to high school, students in the Metropolitan Area of Mexico City who desire to access the public education system to continue their studies are required to take a placement exam designed by the COMIPEMS. The COMIPEMS is composed of nine schools systems that comprise around 600 school-programs.[4,5] Among the ten systems, two are considered elite, as their high schools are affiliated with national universities, as is the case of the UNAM and IPN.[6]

UNAM and IPN are renowned for their elite high school programs, distinguished by superior educational infrastructure and faculty. These institutions boast state of the art laboratories that provide students with the necessary resources to enhance their learning experience. Additionally, facilities such as modern gymnasiums, sports courts, and well-equipped classrooms create an environment conducive to academic learning and extracurricular activities. UNAM and IPN staff have superior credentials compared to other educational systems. Table 1 shows the proportion of teachers holding bachelor's, master's and PhD's degrees by

[4]Excluding UNAM and IPN the school systems are: Colegio de Bachilleres (COLBACH), Colegio Nacional de Educación Profesional Técnica (CONALEP), Dirección General del Bachillerato (DGB), Dirección General de Educación Tecnológica Industrial y de Servicios (DGETI), Dirección General de Educación Tecnológica Agropecuaria y Ciencias del Mar (DGETAyCM), Secretaría de Educación del Gobierno del Estado de México (SE), and Universidad Autónoma del Estado de México (UAEM).

[5]A school-program refers to a specific area of specialization within an educational institution. Not all institutions offer such programs. For instance, IPN and UNAM do not provide programs within their schools. Conversely, CONALEP and SE systems offer programs in various disciplines, including administration, mechanics, and electricity, among others.

[6]Elite schools, defined as those affiliated with the IPN and UNAM systems, have been widely used in the literature. Some studies that use this definition include Ortega Hesles (2015), Dustan et al. (2017), Estrada and Gignoux (2017), Estrada (2017), Pariguana and Ortega (2022), Bobba et al. (2023), and Ngo and Dustan (2024).

high school type (general or technological) and school system.[7] Within general high schools, UNAM has a higher percentage of teachers with a master (22%) and PhD degree (3%) compared to other general high schools. Similarly, within technological high schools, IPN distinguishes itself, with 88% of its teachers holding advanced degrees, compared to lower shares in other systems.

Beyond infrastructure and faculty, UNAM and IPN high schools offer easy entry opportunities to access to their respective affiliated universities. UNAM uses the *pase reglamentado* to give priority admission to students from its affiliated high schools.[8] For the 2015-16 academic year, all 26,611 graduates who applied to UNAM were accepted. In contrast, of the 217,663 applicants who took UNAM university entrance exam, only 11% were admitted (UNAM, 2023). Families perceive enrollment in UNAM affiliated high schools as a strategic approach to gain easy access to UNAM (Chen and Pereyra, 2019). Consequently, some students, despite receiving admission to other high school, opt to take a year off to retake the COMIPEMS exam. The IPN does not provide preferential admission treatment for their high school students. However, for the academic year 2021-22, approximately 60% of students from IPN affiliated high schools were admitted to IPN University, in contrast to only 20% of applicants from outside the IPN system (IPN, 2022 and Instituto Kepler, 2023).

Table 1: Educational Qualifications of Teachers Across School Systems (2015)

System	Bachelor (%)	Master (%)	PhD (%)	Rest (%)	Teachers with degrees known	Total teachers
General High School						
COLBACH	0.92	0.01	0.00	0.07	2,717	2,925
DGB	0.79	0.05	0.02	0.14	145	169
UNAM	0.73	0.22	0.03	0.02	6,212	6,365
SE[1]	0.74	0.11	0.00	0.15	3,172	3,743
UAEM	0.78	0.07	0.00	0.15	686	804
Technological						
DGETA	0.60	0.08	0.05	0.27	46	63
DGETI	0.65	0.07	0.00	0.28	1,570	2,188
IPN	0.83	0.05	0.00	0.12	2,839	3,242
SE[2]	0.71	0.11	0.00	0.18	1228	1,491

Note: This table shows the proportion of teachers holding an advanced degree. The data to produced this table was compiled using statistics from the Survey 911, conducted by the Mexican Ministry of Education.
[1] Includes general high schools under the system of the Ministry of the State of Mexico.
[2] Includes Technological high schools under the system of the Ministry of the State of Mexico.

[7] In the metropolitan area of Mexico City, high schools can be categorized into three distinct types: general, technological, and vocational. General high schools are designed to equip students for higher education. Technological high schools provide students with both a technical degree and the necessary preparation to pursue further education, and vocational high schools focus on preparing students to obtain a technical degree.

[8] *Pase reglamentado* is a policy that provides admission into UNAM without requiring an entrance examination, contingent upon their performance on high school education. Priority is given to candidates who have completed the UNAM high schools within three years and have attained a minimum GPA of 9.0 out of 10.0. Subsequent positions are allocated for students who possess a GPA higher than 7.0 and have concluded their high school education within a span of four years, subject to the availability of slots.

The COMIPEMS exam serves as the only criterion to assign students to schools. Using a serial dictatorship algorithm, the COMIPEMS allocates students to schools by ordering them from the highest to the lowest scoring and matches them based on their rank order list (ROL), that is, the ordered list of preferred school-programs submitted in advance, subject to seat availability. The allocation mechanism produces cutoffs, defined as the test score of the last admitted student, for each school-program, where students meeting or exceeding the cutoff are eligible for admission to the desired program. While the seat allocation mechanism do not change, elite schools, apart of meeting the exam cutoff, require their students to have a middle school GPA of at least seven. Thus, students with GPAs below seven are ineligible for admission to the elite system, even if their placement exam scores are equal or higher to the school-program cutoff.

The COMIPEMS application process begins six months before the high school entrance exam. Ninth graders receive a booklet containing relevant information about the admission process, including details about high school programs characteristics and three years of prior school-programs cutoffs. A month later, after receiving the booklet, participating students complete a registration form that includes questions about their behavior, sociodemographic characteristics, and the goods and services they possess. Students are also required to submit a ROL of up to 20 high schools, ranked from the most preferred to the least preferred. Students take the COMIPEMS exam in June at the end of the school year. The COMIPEMS exam consists of several sections that assess subject-specific knowledge, with a total maximum score of 128 points.[9]

Due to the timing of the admission process, in their first application, students remain unaware of their final middle school GPA at the time of registration. As a result, the majority of students to the left of the GPA threshold imposed by elite high schools optimistically include these institutions in their ROL, anticipating that they would achieve a GPA of at least seven, thereby meeting the eligibility criterion. Upon middle school GPA realization, students with GPAs lower than seven adjust their ROL submissions in subsequent applications not listing elite schools. Table 2 presents the proportion of retakers who include an elite school in their ROL by GPA for retakers in their first and second applications. In 2015, 84% and 90% of the retakers, with GPAs slightly below the threshold (6.9 GPA) and at the threshold (7.0 GPA), included at least one elite school in their ROL during their initial attempt at the COMIPEMS examination. However, in their next attempt, these proportions changed to 1% and 86%, demonstrating how students adjust their ROL upon knowing their

[9]The COMIPEMS evaluates the following areas: Verbal Skills (16), Spanish (12), History (12), Geography (12), Civic and Ethical Education (12), Mathematical Skills (16), Mathematics (12), Physics (12), Chemistry (12), and Biology (12). The global score is the sum of these components.

realized GPA.

Figure 1: COMIPEMS PROCESS

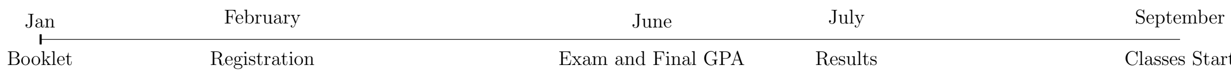


Table 2: Proportion of Students Ranking at least one Elite School

	All students 2015[1]				Retakers first take 2015			Retakers second take 2016			
GPA middle school	**First**	**top 5**	**All**	**N**	**First**	**top 5**	**All**	**First**	**top 5**	**All**	**N**
6.5	0.38	0.45	0.51	2,025	0.58	0.65	0.71	0.01	0.01	0.01	180
6.6	0.42	0.50	0.55	3,284	0.58	0.66	0.71	0.01	0.01	0.01	304
6.7	0.46	0.53	0.59	4,686	0.65	0.72	0.79	0.01	0.01	0.01	351
6.8	0.48	0.55	0.61	6,051	0.63	0.71	0.77	0.00	0.00	0.00	550
6.9	0.51	0.59	0.64	7,357	0.70	0.79	0.84	0.01	0.01	0.01	586
7.0	0.58	0.68	0.74	8,068	0.79	0.86	0.90	0.75	0.82	0.86	985
7.1	0.58	0.67	0.73	8,809	0.80	0.87	0.90	0.77	0.85	0.88	1,023
7.2	0.60	0.68	0.73	9,157	0.80	0.87	0.91	0.78	0.83	0.86	1,089
7.3	0.60	0.69	0.75	9,668	0.81	0.87	0.90	0.79	0.84	0.87	1,115
7.4	0.61	0.69	0.74	9,891	0.82	0.86	0.89	0.80	0.85	0.87	1,053
7.5	0.62	0.69	0.74	9,845	0.84	0.88	0.91	0.80	0.85	0.87	1,153
7.6	0.62	0.69	0.74	9,825	0.86	0.90	0.92	0.80	0.86	0.88	1,103
7.7	0.62	0.70	0.75	9,864	0.87	0.90	0.92	0.83	0.86	0.89	1,052
7.8	0.64	0.72	0.76	9,499	0.88	0.91	0.93	0.83	0.88	0.90	1,036
7.9	0.64	0.71	0.76	9,190	0.89	0.92	0.93	0.87	0.90	0.92	1,014
8.0	0.65	0.73	0.77	9,312	0.90	0.94	0.95	0.87	0.91	0.92	1,066

Note: The proportions reflect the share of students who listed at least one elite school in their preference list, either as their first choice, within the top five, or anywhere in the list.
[1] All students 2015 population excludes retakers from 2014.

Elite schools have higher demand compared to non-elite institutions. The ratio of students ranking IPN and UNAM as their first option by the total number of seats offered is 2.5 and 5, respectively. These figures imply that even if every student who selected IPN or UNAM as their first choice were admitted, there would still be a deficit of 1.5 seats for IPN and 4 seats for UNAM per admitted student. The substantial over-subscription on elite schools reflects the perceived value and the competitive pressure associated with securing admission to these institutions. For non-elite institutions, the applicant to seat ratio remains relatively low, with a maximum hovering around one (Figure 2). The high demand for elite programs results in higher admission cutoff scores, which in turn reflects the competitive environment for elite schools seats, as the cutoffs are determined by the test score of the last admitted student. Schools with high cutoffs are at full capacity, while schools with low cutoffs may still have available seats (Figure 3.)

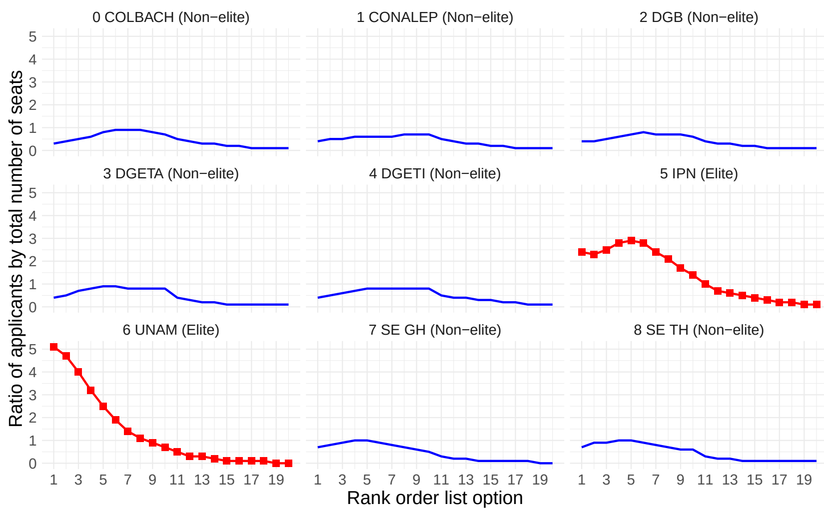


Figure 2: Over-subscription

Notes: This figure displays the ratio of applicants to available seats across school options in the COMIPEMS system, by ROL position. The ratio is calculated as the number of students ranking a school as their k-th option divided by the number of seats available at that school.

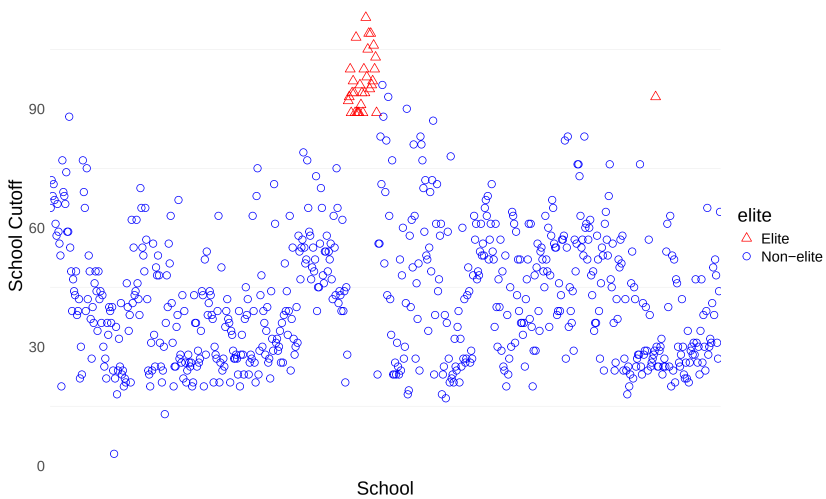


Figure 3: School-program cutoffs

Notes: This figure shows the cutoff (the test score of the last admitted student) for each school-program in 2015. Each point represents a cutoff score under the serial dictatorship algorithm. Triangles indicate elite schools (UNAM and IPN), while circles denote non-elite institutions.

2.1 Policy Rule

Although neither UNAM nor IPN have officially stated the rationale behind the minimum GPA policy, it is likely that their admission criterion is driven by a desire to maintain academic standards, ensure that incoming students are adequately prepared to succeed in their programs, and manage the high demand for the limited number of seats offered in their schools. With a significant number of applicants competing for a relatively small number of seats, setting higher GPA requirements helps these institutions select candidates who are most likely to excel and benefit from their academic offerings. In particular, when accepting students who are not sufficiently prepared may have negative consequences in graduation rates (Denning et al., 2022 and Arcidiacono et al., 2016).

The policy rule is sharp, no student with a GPA lower than seven is accepted into the elite system. Figure 4 illustrates the effects of the policy in students school allocation using histograms of raw test scores in the entrance exam in four groups, in the 2016 COMIPEMS process, formed by the interaction of elite and non-elite schools and first-time takers and retakers. The vertical line in Figure 4 shows the minimum test score required for a student to gain admission to an elite high school. Students with test scores to the right of the line meet the admission requirement. Exclusively students with a GPA of seven or higher are matched with elite schools (left histograms). Conversely, all students with GPAs below seven are matched with non-elite schools (right histograms). The policy is binding for students with GPAs below 7.0 with sufficient test scores to qualify for elite schools.

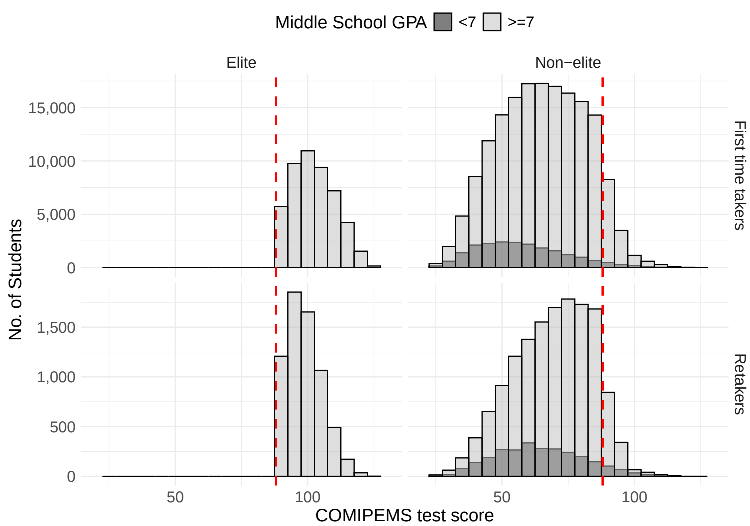


Figure 4: Sharp Policy Cutoff

Note: This figure shows histograms of raw COMIPEMS scores in 2016 for four groups (elite vs. non-elite $\times$ first-time takers vs. retakers). The vertical line in the left panel is positioned at the minimum score required for elite admission.

2.2 Retakers

Retaking the high school entrance exam is costly. Students must invest time in studying for the test again, often while attending school, and may incur monetary costs associated with preparatory courses or materials. For those already enrolled in a school, transferring entails further costs, such as adapting to a new environment, potentially repeating coursework, and in some cases, restarting high school entirely due to non-transferable credits. Despite the associated costs, around 8.5% of students in the 2015 cohort chose to retake the exam in 2016.

Numerous factors contribute to the decision to retake, including dissatisfaction with the initial assignment, family relocation, or setbacks such as failing a grade or being expelled from school. These factors are not expected to change abruptly for students just above or below the GPA threshold. However, at the 7.0 middle school GPA threshold, which determines eligibility for elite schools, retake incentives are shaped by competing forces. On one hand, students may have already been matched with an elite school in their first attempt, diminishing the retake probability. On the other hand, students with GPAs at or above 7.0 may be more likely to retake because they become eligible for elite schools and thus stand to gain from a better assignment. As a result, a discontinuity in retake behavior may emerge at the threshold, reflecting the net effect of these opposing mechanisms.

These competing incentives give rise to a selection concern. More motivated students may be inclined to retake the exam if they were not matched with their preferred elite school, aiming to improve their assignment. In contrast, the most talented students are less likely to appear among retakers, as they likely secured admission to elite schools on their first attempt. Additionally, motivated students with a GPA below 7.0 may choose not to retake, since the policy restricts their options to non-elite schools. Consequently, individuals retaking the examination near the eligibility cutoff for elite schools may exhibit differences in unobservable characteristics such as ability and motivation.

3 Data and Methodology

3.1 Data

The data used in this paper combines the COMIPEMS entrance surveys, standardized test scores, and ROLs for all applicants in 2015 and 2016. The datasets encompass detailed individual-level data, comprising demographic information, socioeconomic background such as parental education, family income, scholarship details, self-reported student behavior, and socioemotional indicators. The study concentrates on individuals retaking the examination, since those undertaking the entrance examination for the first time lack awareness of their middle school GPA policy, and consequently, changes in effort are not anticipated for them. I restrict the analysis to students who retook the exam the following year, excluding multi-year retakers, as their decision to retake may be influenced by different factors than next year retakers. I also exclude students affiliated with The National Institute for Adult Education (INEA, by its acronym in Spanish), who are 18 years or older, when taking the COMIPEMS exam.[10] Out of state students are dropped as well, as they need to reallocate in order to attend school and may be extra motivated.[11] Additionally, I exclude students who did not prioritize an elite school at the top of their ROL, since the assignment mechanism processes preferences sequentially, and listing a non-elite school first effectively removes elite options from consideration. I drop students who initially matched with an elite school, as their motive for retaking is unlikely to reflect a desire to access an elite option, since they previously secured a seat. Finally, I drop students with GPA below 7.0 who have sufficient scores to attend an elite school, to maintain consistency with the exclusion of students above the threshold who actually were enrolled in elite schools. This ensures that the sample excludes all students who could have accessed elite options regardless of their GPA, whether through actual admission or potential eligibility in the absence of the policy.

Table 3 reports the match outcomes for first-time takers, all retakers, and the analyzed sample of retakers by the type of schools students listed in their ROL. ROLs are classified into three categories. Students who exclusively listed elite schools appear under the Elite column; students who listed both elite and non-elite schools appear under Mix; and students who listed only non-elite schools appear under Non-elite. The outcomes are decomposed into Matched, referring to students who were assigned to a school by the serial dictatorship algorithm, and Unmatched, students who were not assigned by the algorithm but are still allowed to choose a school with seat availability, as upper secondary education is compulsory in Mexico.

In 2015, approximately 300,000 students participated in the COMIPEMS high school admissions process. Of these, 274,430 were first-time takers, while 25,058 were retakers.

[10] The INEA is a federal agency in Mexico dedicated to providing basic education services to adults who did not complete their formal schooling. INEA specializes in literacy, primary, and secondary education for individuals aged 15 and older, with flexible learning models tailored to adult learners. Its programs aim to reduce educational gaps and promote social inclusion by offering free educational materials and community-based instruction. For instance, a student starting middle school at 15 years old will graduate at 18 years old.

[11] The are fewer than ten out-of-state students in the retakers population, rendering it improbable that their inclusion would significantly influence the estimations.

The final sample of retakers comprises 20,714 individuals. Retakers disproportionately target programs with higher admission cutoffs: the share applying exclusively to elite schools is 12.3% among first-time takers, compared with 20.6% for all retakers and 23.0% in the final retaker sample. Consistent with these tougher targets, 85.5% of first-time takers were matched through the centralized serial-dictatorship mechanism, whereas only 61.6% of retakers secured a seat. The figure falls further to 58.3% for the restricted retaker subsample (those who prioritized an elite school and met additional criteria).

Table 3: Breakdown of Match Outcomes by ROL Type

Panel A: School Placement 2015, First-time Takers				
ROL Type	Elite	Mix	Non-elite	Total
Matched	18,284 (6.7%)	160,558 (58.5%)	55,909 (20.4%)	234,751 (85.5%)
Unmatched	15,348 (5.6%)	19,766 (7.2%)	4,565 (1.7%)	39,679 (14.5%)
Total	33,632 (12.3%)	180,324 (65.7%)	60,474 (22.0%)	274,430 (100.0%)
Panel B: School Placement 2015, Retakers				
All Retakers				
Matched	145 (0.6%)	13,815 (55.1%)	1,464 (5.8%)	15,424 (61.6%)
Unmatched	5,013 (20.0%)	4,343 (17.3%)	278 (1.1%)	9,634 (38.4%)
Total	5,158 (20.6%)	18,158 (72.5%)	1,742 (7.0%)	25,058 (100.0%)
Subpopulation of Retakers				
Matched	0 (0.0%)	12,069 (58.3%)	0 (0.0%)	12,069 (58.3%)
Unmatched	4,768 (23.0%)	3,877 (18.7%)	0 (0.0%)	8,645 (41.7%)
Total	4,768 (23.0%)	15,946 (77.0%)	0 (0.0%)	20,714 (100.0%)
Panel C: School Placement 2016, Retakers				
All Retakers				
Matched	3,053 (12.2%)	11,541 (46.1%)	3,313 (13.2%)	17,907 (71.5%)
Unmatched	4,465 (17.8%)	2,140 (8.5%)	546 (2.2%)	7,151 (28.5%)
Total	7,518 (30.0%)	13,681 (54.6%)	3,859 (15.4%)	25,058 (100.0%)
Subpopulation of Retakers				
Matched	2,734 (13.2%)	10,096 (48.7%)	1,606 (7.8%)	14,436 (69.7%)
Unmatched	4,130 (19.9%)	1,855 (9.0%)	293 (1.4%)	6,278 (30.3%)
Total	6,864 (33.1%)	11,951 (57.7%)	1,899 (9.2%)	20,714 (100.0%)

Notes: Matched refers to students assigned by the serial dictatorship algorithm using their rank-order lists (ROLs). *Unmatched* denotes students not assigned by the algorithm who retain the right to enroll in another school provided seats are available, as upper-secondary education is compulsory in Mexico. The *Subpopulation of Retakers* consists of students who ranked an elite school first, did not match to an elite school initially, and satisfy the restrictions in Section 3.1. *ROL Type* classifies each student's ROL into three categories: *Elite* if the list contains only elite schools; *Mix* if it contains at least one elite and one non-elite school; and *Non-elite* if it contains only non-elite schools. Percentages in parentheses are shares of the panel total within each subtable.

3.2 Methodology

The objective of this study is to measure the change in effort exerted from having the opportunity to enroll in elite schools measured by difference in test scores between students with the opportunity to attend elite schools and those prevented from enrolling into them. Let T be a binary treatment indicator, where $T = 1$ if a student is eligible for admission into an elite high school and $T = 0$ otherwise. Let R denote the running variable which in this context is the student's middle school GPA. At the known cutoff $r_0 = 7$, the probability of eligibility jumps from zero to one. Let Y^* denote the true outcome of interest, which is only observed for a subset of individuals selected through a non-random process. Define Y as the observed outcome and S as a binary indicator for sample inclusion, where $S = 1$ if the individual is included in the sample and $S = 0$ otherwise. Therefore, $Y = Y^*$ when $S = 1$, and Y is missing when $S = 0$. In this setting, S indicates whether a student chooses to retake the COMIPEMS exam. Thus, $S = S_1T + S_0(1 - T)$ gives the potential sample selection under treatment or no treatment. In addition, there is information about predetermined covariates denoted by W.

Due to the strict policy in place, all students are compliers. Thus, students possessing a GPA under seven are categorically precluded from accessing elite schools. Exploiting the discontinuity in the middle school GPA, the aim is to recover the local average treatment effect given the data on $\{Y, S, T, R, W\}$. Formally, the model is:

$$
\begin{aligned}
Y^* &= \tau_O T + W\delta_1 + U && (1)\\
T &= \mathbf{1}(R \geq r_0) && (2)\\
R &= W\delta_2 + V && (3)\\
S^* &= \tau_S T + W\delta_3 + E, \quad S = 1(S^* > 0) && (4)\\
Y &= Y^* \cdot S && (5)
\end{aligned}
$$

where U, V and E are error terms, δ_1, δ_2 and δ_3 are coefficients associated with W, and τ_O and τ_S are the effect of treatment T on the outcome variable and the selection term. Assuming no sample selection $\lim_{r \downarrow r} E[S \mid R = r] = \lim_{r \uparrow r_0} E[S \mid R = r]$, that is $\tau_S = 0$, and under the assumptions of the canonical RD design

A1. Discontinuity: There is a discontinuity in treatment at the cutoff.

$$\lim_{r \downarrow r_0} E[T \mid R = r] \neq \lim_{r \uparrow r_0} E[T \mid R = r].$$

A2. Smoothness: The conditional expectations of the potential outcomes are continuous at the cutoff:

$$\lim_{r \downarrow r_0} E[Y_t^* \mid R = r] = \lim_{r \uparrow r_0} E[Y_t^* \mid R = r], \quad \text{for } t \in \{0, 1\}.$$

the LATE on the outcome variable can be recovered using the observed sample as:

$$\begin{aligned}
\lim_{r \downarrow r_0} E[Y \mid R = r] - \lim_{r \uparrow r_0} E[Y \mid R = r] = \tau_O + & \\
& \lim_{r \downarrow r_0} \sum_{w,u} (w\delta_1 + u) \Pr[W = w, U = u \mid R = r] - \\
& \lim_{r \uparrow r_0} \sum_{w,u} (w\delta_1 + u) \Pr[W = w, U = u \mid R = r] \\
= \tau_O &
\end{aligned}$$

The model, without selection, resembles a conventional endogenous variable configuration, but with an observed assignment variable R. Moreover, W can be determined endogenously, provided that it is realized before V, so there is randomization at the cutoff r_0. Intuitively, students do not have full control over their GPAs. Within this framework, the variation in outcomes among individuals is entirely characterized by the random variables (W, U). Consequently, individuals who exhibit identical values of (W, U) will experience one of two possible outcomes, contingent upon receiving treatment. There is no stance as to whether some elements of δ_1 or δ_2 are zero, since there is not need for exclusion restriction under the RD design.

However, if there is sample selection $\lim_{r \downarrow r_0} E[S \mid R = r] \neq \lim_{r \uparrow r_0} E[S \mid R = r]$, that is $\tau_S \neq 0$ the LATE cannot be estimated using the observed sample. That is $\tau = \lim_{r \downarrow r_0} E[Y^* \mid R = r] - \lim_{r \uparrow r_0} E[Y^* \mid R = r] \neq \lim_{r \downarrow r_0} E[Y \mid R = r] - \lim_{r \uparrow r_0} E[Y \mid R = r]$. Given that $\lim_{r \downarrow r_0} \sum_{w,u} (w\delta_1 + u) \Pr[W = w, U = u \mid R = r] \neq \lim_{r \uparrow r_0} \sum_{w,u} (w\delta_1 + u) \Pr[W = w, U = u \mid R = r]$. Even if the distribution of observables W is smooth at the cutoff, discontinuities in the selection process can still generate differences in the distribution of unobservables U. In other words, balance in predetermined covariates does not guarantee valid identification if the probability of being observed changes at the cutoff.

The credibility of $\lim_{r\downarrow r_0} E[S \mid R = r] = \lim_{r\uparrow r_0} E[S \mid R = r]$ is challenged in this context due to self-selection.[12] Specifically, I only observe outcomes for students who opt to retake the COMIPEMS exam, with this decision being endogenous. Such endogeneity may result in differences between the student populations on either side of the cutoffs. Methodologies dealing with sample selection employ some correction (e.g., Heckman, 1979 and Heckman, 1990) that relies on an exclusion restriction, which may be questionable in the RD design context, when there is a jump in both the selection and the outcome variables at the running variable discontinuity.

To address self-selection in our empirical context, I implement a set of alternative econometric approaches under the RD design umbrella, summarized in Table 4. These approaches differ in their identification assumptions and target estimands. Under smoothness condition, apart from the canonical RD design, I take two approaches to deal with sample selection. In the first approach I subtract from the outcome variable its lagged value (Lee and Lemieux, 2010) to recover smoothness, while in the second I explicitly account for sample selection by assuming smoothness in both the outcome and selection variables (Dong, 2019). I also consider approaches that depart from the smoothness condition. These alternative methods rely on the assumption of unconfoundedness, enabling the estimation of conditional average treatment effects (CATE) either through adjustment for observable characteristics using machine learning techniques (Athey et al., 2019), or by accounting for unobserved characteristics via a proxy variable (Dale and Krueger, 2014). An exposition of each model follows in detail.

Gains model. The gains model estimates the LATE using the difference in test scores as the outcome variable. This approach is based on the insight that if the lagged outcome is determined prior to treatment assignment (i.e., before the realization of the running variable R), then subtracting the lagged outcome from the current outcome, $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$, restores the smoothness condition required for identification (Lee and Lemieux, 2010). However, this strategy may be problematic if students engage in anticipatory behavior, adjusting their effort in response to expected elite school eligibility based on prior academic performance,

[12] A general approach to estimate sample selection is using a fuzzy RD Wald ratio:

$$\tau_S = E[S_1 - S_0 \mid R = r_0, C] = \frac{\lim_{r\downarrow r_0} E[S \mid R = r] - \lim_{r\uparrow r_0} E[S \mid R = r]}{\lim_{r\downarrow r_0} E[T \mid R = r] - \lim_{r\uparrow r_0} E[T \mid R = r]}.$$

In this case, treatment T switches deterministically at the cutoff, middle school GPA equal to 7.0. The denominator equals one and the estimand reduces to a sharp RD design.

Table 4: Summary of RDD Models, Conditions, and Effect Estimates

RD under smoothness	**Continuity**[1]	**Other Condition**	**Estimated Effect**	**Assumption Source**
Canonical	$E[Y_t^* \mid r]$	Continuity $E[S_t \mid r_0]$	LATE	Hahn et al. (2001)
Gains	$E[\Delta Y_t \mid r]$	No anticipatory behavior	LATE	Lee and Lemieux (2010)
Sample Selection	$E[(Y_t^*, S_t) \mid r]$	—	LATE and Bounds	Dong (2019)
RD under unconfoundedness				
Observables	—	Unconfoundedness	CATE	Athey et al. (2019)
Unobservables	—	Unconfoundedness	CATE	Dale and Krueger (2014)

[1] Continuity refers to the assumption that the conditional expectation of potential outcomes is continuous at the cut-off r_0 of the running variable. For instance, for the canonical RD design, $E[(Y_t^*) \mid r]$ implies $\lim_{r \downarrow r_0} E[Y_0^* \mid R = r_0] = \lim_{r \uparrow r_0} E[Y_1^* \mid R = r_0]$.

since this would render the lagged outcome endogenous and lead to biased estimates.[13] As noted in prior literature, the lagged-score specification tends to outperform the gain-score approach across a range of value added estimation settings (Andrabi et al., 2011; Guarino et al., 2015; McCaffrey et al., 2009). In the results Section 4, I include both gains and lagged specifications due to their conceptual similarity. However, it is only the former that reestablishes the smoothness condition.

Sample Selection model. The model deals with both treatment endogeneity and sample selection (Dong, 2019). The identification relies on smoothness $E(Y_t^*, S_t|r)$ condition for the entire population and does not require specifying any selection mechanism. Under this assumption the observed sample selection $S = S_0 + T(S_1 - S_0)$ is allowed to jump at the cutoff r_0. The model separates the policy impact into two components the extensive margin, reflecting the change in the retaking probability, and the intensive margin, capturing the changes in test scores. Formally, the extensive margin effect is defined as $E[S_1 - S_0 \mid R = r_0]$ and the intensive margin effect is expressed as $E[Y_1^* \mid S_1 = 1, R = r_0] - E[Y_0^* \mid S_0 = 1, R = r_0]$.

While the intensive margin effect identifies the change in outcomes for students who retake the exam, this estimate may still be confounded by selection into retaking. To address this issue, bounds on the intensive margin can be constructed under the assumption of monotonic selection, that requires that treatment can only affect sample selection in one direction. In this case, students with GPAs lower than seven are less likely to retake the exam $Pr(S_1 \geq S_0) = 1$. These bounds are sharp and fall into the class of *worst case* bounds scenario (Horowitz and Manski, 1995). In our data, monotonic selection is expected to hold

[13] In this context, anticipatory effort exerted in earlier grades in expectation of eligibility may invalidate the assumption that the lagged outcome is predetermined.

for students with insufficient test scores to be matched to an elite school in their initial application.[14]

Departing from the RD design setting, the next models relax the continuity assumption and instead rely on unconfoundedness $Y_0, Y_1 \perp T \mid W$. Conditional on the covariates W, the treatment assignment is as good as random. Under this assumption, I use the covariate adjusted difference between the two treatment groups, $E[Y_1 \mid T = 1, W] - E[Y_0 \mid T = 0, W]$ to estimate the Conditional Average Treatment Effect (CATE). Without assuming continuity of potential outcomes in the running variable or the absence of endogenous sorting at the cutoff, this framework does not recover a LATE at the threshold. Instead, it recovers a average treatment effect. As such, the estimand under unconfoundedness assumption lacks the clear quasi-experimental interpretation associated with an RD design. Nevertheless, by leveraging rich covariate information, this approach offers a flexible and policy-relevant alternative when the assumptions of traditional RD design may be violated.

Model based on observables. Given the high dimensionality and richness of the available covariates, I implement random forests to estimate the conditional average treatment effect (CATE). The estimator is not valid without strong assumptions (Angrist and Rokkanen, 2015; Peng and Ning, 2021). The model controls for a comprehensive set of sociodemographic variables (e.g., income, mother's education, access to goods and services), behavioral variables (e.g., prioritizing activities, meeting deadlines), and psychological variables (e.g., feelings of depression, aggressiveness). The model is estimated using a linear forest approach (Athey et al., 2019).

Model based on unobservables. This model controls for the unobserved ability of students, since a positive bias may arise because more ambitious or motivated students, with higher unobserved ability, are more likely to opt into retaking. By controlling for unobserved ability through a proxy variable, such as the proportion of elite schools in the student's rank-order list, I address the selection problem that arises from omitted variables. This strategy relies on the idea that students' revealed preferences embed information about latent traits such as ambition, expectations, or private knowledge of ability. Similar to other settings where admission processes are leveraged to address selection on unobservables, incorporating preference data allows us to proxy for these latent characteristics and mitigate bias in the estimated effects (Dale and Krueger, 2014).

[14]Under monotonic selection, with probability one, selection into retaking under eligibility for elite schools is at least as high than under non-eligibility. This holds after excluding students who were matched to an elite school on their first attempt in 2015, as these students are empirically less likely to retake (see Table 5).

3.2.1 Estimation

Estimations rely on first-order local linear specification Calonico et al. (2014). The treatment effect at the cutoff is obtained by fitting separate weighted regressions on either side of the running variable threshold and taking the difference in intercepts. A triangular kernel is used to place greater weight on observations closer to the cutoff, and bandwidths are selected using mean-squared error (MSE)–optimal rules with robust bias correction to ensure valid inference. Higher-order local polynomials are not employed, as they often generate noisy estimates and result in poor confidence interval coverage (Gelman and Imbens, 2019; Kolesár and Rothe, 2018).

Following Calonico et al. (2014), τ_{RD} is estimated using local polynomial regression with kernel weights for a generic outcome $\tilde{Y}$ (e.g., Y or S). For each side of the cutoff r_0, the local linear estimator solves[15]

$$\hat{\beta}_+(h) = \arg\min_{\alpha,\beta} \sum_{i:R_i \geq r_0} \Big(\tilde{Y}_i - \alpha - \beta(R_i - r_0)\Big)^2 K_h(R_i - r_0),$$

$$\hat{\beta}_-(h) = \arg\min_{\alpha,\beta} \sum_{i:R_i < r_0} \Big(\tilde{Y}_i - \alpha - \beta(R_i - r_0)\Big)^2 K_h(R_i - r_0),$$

where $K_h(u) = K(u/h)/h$ is a rescaled kernel function with bandwidth h and $K(\cdot)$ is the triangular kernel. Due to recentering the fitted values at the cutoff are given by

$$\hat{\mu}_+(r_0) = \hat{\alpha}_+, \qquad \hat{\mu}_-(r_0) = \hat{\alpha}_-,$$

and the RD estimator is

$$\hat{\tau}_{\mathrm{RD}} = \hat{\mu}_+(r_0) - \hat{\mu}_-(r_0).$$

Because local polynomial estimators at boundaries are biased, Calonico et al., 2014 propose a bias-corrected estimator. The leading bias is estimated using a higher-order polynomial for both estimations.

$$\hat{\tau}_{\mathrm{BC}} = \hat{\tau}_{\mathrm{RD}} - \hat{B},$$

where $\hat{B}$ is the estimated bias. Robust inference is then based on

$$CI_{\mathrm{RBC}}: \quad \hat{\tau}_{\mathrm{BC}} \ \pm \ z_{1-\alpha/2} \cdot \widehat{\mathrm{SE}}_{\mathrm{RBC}}.$$

[15]Covariates are added for estimations based on observables and unobservables.

Bandwidths h_-, h_+ are selected using mean squared error (MSE), optimal rules for point estimation, while coverage, optimal bandwidths are used for confidence intervals. In this setting, the running variable is discrete. Therefore, nearest-neighbor variance estimators are employed to ensure valid inference and correct coverage of confidence intervals.

I re-estimate the models by dropping students with middle school GPA of seven to control for potential heaping using the donut-RD design approach (Almond et al., 2010). Heaping might occur if some students are able to negotiate a GPA of seven, even when their grades were lower. The main reason heaping is not expected is that the GPA in middle school is the average of the grades obtained in each subject in middle school. Each teacher in middle school submits their grades individually, and then these grades are averaged by a system. For a student to manipulate the system, the average GPA must be consistent with the entire set of grades submitted.[16] To do so, a student would likely need to convince several teachers and possibly even the principal to rig the system. While I do not entirely dismiss the possibility of such manipulation, I believe it is unlikely to occur. Nevertheless, I estimate the regression by excluding students who are exactly at the seven cutoff.

4 Results

This section first shows how the middle school GPA policy induces self-selection. It then presents the results on effort, measured as the difference in test scores between students deemed eligible and those deemed ineligible for admission to elite institutions.

4.1 Self-Selection

The middle school GPA policy for elite eligibility operates through two channels with respect to retaking: (i) among students eligible to attend elite schools, those who are initially matched to an elite school exhibit a lower probability of retaking the exam; and (ii) among students who are eligible but not initially admitted, the policy increases the probability of retaking relative to otherwise similar students who are not eligible. Taken together, these patterns show a discontinuous change in selection at the relevant middle school GPA cutoff $\lim_{r\downarrow r_0} E[S \mid R = r] \neq \lim_{r\uparrow r_0} E[S \mid R = r]$. The policy can be expected to generate monotonic selection for students with insufficient test scores to be matched in their first attempt to elite schools. These students, under treatment are weakly more likely to retake the exam ($Pr(S_1 \geq Pr(S_0))$ than without treatment. The monotonicity condition is necessary to build

[16] Appendix A1 contains detailed information about middle school grading system.

sharp bounds of the effort LATE (Dong, 2019).[17]

4.1.1 First-Round Admission and the Probability of Retaking

To assess whether first-round elite admission affects the probability of retaking, I define the *distance to least-preferred elite school*. For each student i, let the ROL be

$$\mathrm{ROL}_i \;=\; (s_{i1}, s_{i2}, \ldots, s_{iJ}), \quad s_{i1} \succ s_{i2} \succ \cdots \succ s_{iJ},$$

where s_{i1} is most preferred. Let $\mathcal{E}$ denote the set of elite schools, and for any school s let c_s be its first-round admission cutoff, the test score of the last admitted applicant. Define the student's least-preferred elite option as

$$s_i^L \;\equiv\; \arg\min_{\succ}\{\, s_{ij} \in \mathrm{ROL}_i : s_{ij} \in \mathcal{E} \,\},$$

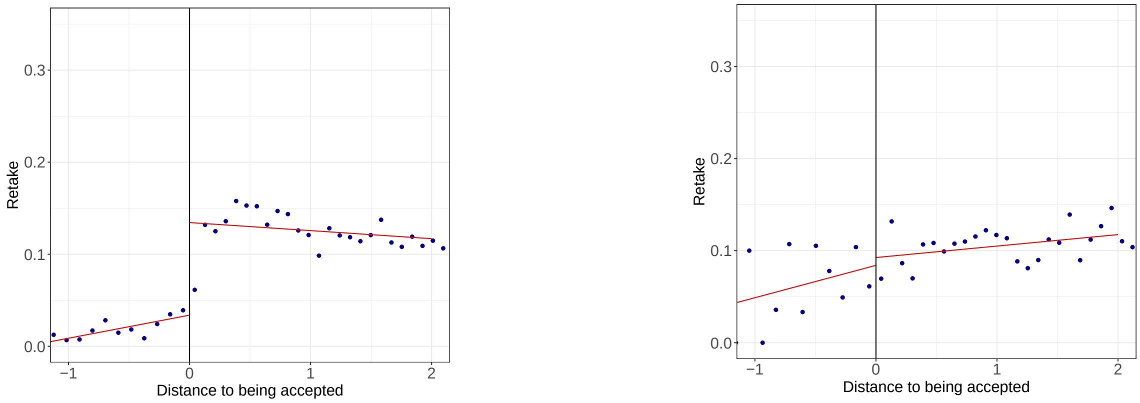
i.e., the lowest-ranked elite school on i's list, and set $c_i^L \equiv c_{s_i^L}$. The variable of interest is

$$\mathrm{DLP}_i \;\equiv\; c_i^L - TS_i,$$

where TS_i is i's entrance-exam score. Thus, $\mathrm{DLP}_i \leq 0$ indicates TS_i meets or exceeds the cutoff for the least-preferred elite), whereas $\mathrm{DLP}_i > 0$ indicates the student fell short of all listed elite cutoffs. Students with $\mathrm{GPA}_i < 7$ remain ineligible for elite admission regardless of DLP_i.

I check for discontinuities in the retake probability at 0. Figure 5a shows the relationship between retake probability and the *distance to least-preferred elite school* variable for students eligible to elite school. There is a discontinuity in retake at 0, which can be interpreted as students admitted to elite schools being unlikely to retake the COMIPEMs exam. In contrast, Figure 5b, which presents the same relationship for students with GPAs below 7.0, reveals no discontinuity in retake behavior around the cutoff. This suggests that, for students below the GPA threshold, retake behavior is influenced by factors unrelated to elite school admission, given their ineligibility under the policy.

[17] The monotonic condition aligns with the sample construction, for details see subsection 3.1.



(a) Students with GPA ≥ 7 (b) Students with GPA < 7

Figure 5: Retake Probability by Distance to Least-Preferred Elite School

Notes: This figure plots the probability of retaking in 2016 against the *distance to least-preferred elite school* variable. A negative *distance to least-preferred elite school* indicates a test score that is sufficient for elite school admission. A positive *distance to least-preferred elite school* indicates a test score insufficient for elite school admission. Panel (a) focuses on students with $GPAs \geq 7$. Panel (b) presents results for students with $GPAs < 7$.

To formally estimate the impact of having a test score equal to or greater than the admission school cutoff, I estimate a RD separately for students eligible for elite schools and those ineligible. Table 5 reports the results for eligible and ineligible students. In 2015, admission to an elite school reduces the probability of retaking by 2.9 percentage points, while no effect is observed for ineligible students, as expected. This is consistent with the notion that highly motivated or talented students are already matched to elite schools and thus are less likely to appear in the sample.

Table 5: LATE of Initial Elite Matching on Retake Probability

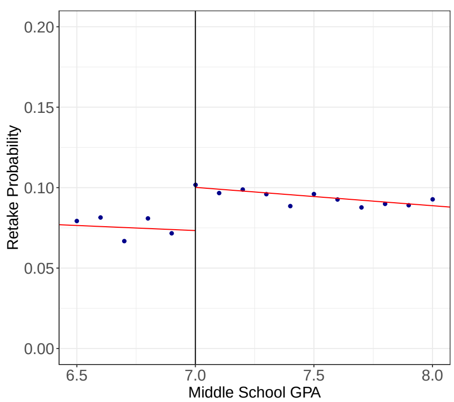
	Eligible (GPA $\geq$ 7.0)	Ineligible (GPA $<$ 7.0)
τ_S	0.029**	0.011
SE	0.012	0.037
Bandwidth (h)	0.33	0.53
Bias (L - R)	0.001 − 0.004	0.001 − 0
Obs. (L - R)	1,855 − 2,971	309 − 812

Notes: The table reports the LATE of having a sufficient test score to be admitted into a elite school on retake probability. Standard errors (SE) are robust. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

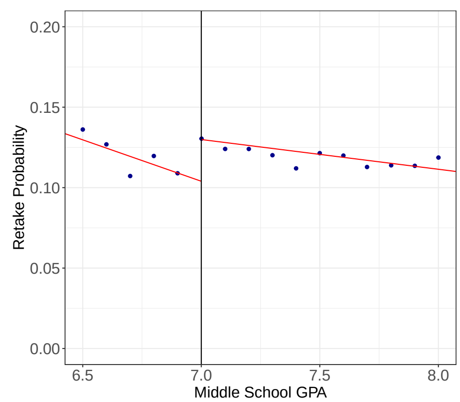
4.1.2 Probability of Retaking and Middle School GPA

To assess whether eligibility for elite schools based on middle school GPA affects the probability of retaking the entrance exam, I explore retake behavior around the 7.0 GPA threshold.

Figure 6a shows that the probability of retaking the COMIPEMS exam exhibits a clear discontinuity at the 7.0 GPA threshold for all retakers. At this cutoff, approximately 10% of students retake the exam, compared to 7.5% at a 6.9 GPA, representing a 25% decrease in retake probability between students with access to elite schools and those without. The discontinuity reflects the combined influence of two opposing mechanisms: students matched with elite schools on their first attempt are less likely to retake, while those who are eligible but were not matched may choose to retake with the intention of securing seat at an elite school. To isolate the latter effect, Figure 6b drops students with sufficient test scores to be matched with elite schools, that is with $DLP_i \leq 0$. The discontinuity persists, with close to 12.5% of students at 7.0 retaking the exam, compared to 10% at 6.9, a 20% decrease, capturing the incentive to gain access to the elite school system.



(a) All students



(b) Students with $DLP_i > 0$

Figure 6: Regression Discontinuity Results for 2016

Notes: This figure shows regression discontinuity results relating middle school GPA to the probability of retaking the COMIPEMS exam. Panel A plots all students, while Panel B restricts the sample to retakers who did not achieve sufficient test scores to access elite schools in their first opportunity $DLP > 0$.

To formally quantify the change in retake probability. I estimate a RD, replacing the running variable DLP_i with middle school GPA_i at the 7.0 eligibility threshold. Table 6 reports baseline regression discontinuity estimates of the 7.0 GPA policy on the probability of retaking the COMIPEMS exam in 2016. Column (1) presents results for the full sample of students, indicating an increase in the probability of retaking of 3.0 percentage points. Given a baseline probability of retaking of about 10 percent at the threshold, this effect corresponds to roughly a 30 percent higher probability of retaking for students with the opportunity to attend elite schools. This estimate captures two sources of discontinuity: (i) eligibility for elite schools under the 7.0 GPA policy and (ii) assignment to an elite school in the previous year, which reduces the probability of retaking. Column (2) reports results for the students with $DLP_i > 0$, which excludes all students with sufficient test scores to be matched with elite schools, regardless of whether they were eligible under the GPA policy. In this subsample, the RD design isolates the causal effect of eligibility on retaking without contamination from prior elite school assignment. The students in these sample are the ones with positive values of the variable *distance to least-preferred elite school*. The estimated effect is 3.4 percentage points, which is consistent with the notion that prior assignment to an elite school lowers the incentive to retake. Relative to the same baseline retake probability of 10 percent, this effect implies an increase of 34 percent for students eligible to attend elite schools derived from the middle school GPA policy.

These results show that sample selection is a concern around the cutoff. The discontinuity at the 7.0 GPA threshold impact students' decisions to retake the exam. However, information about the underlying process driving students choices is lacking. Although some student characteristics can be compared at the middle school GPA cut-off, others, such as motivation or aptitude, are unobservable and therefore cannot be directly tested for students to the right and the left of the middle school GPA threshold. Without data on how students make and act upon the retake decision, I cannot determine whether the composition of test-takers around the threshold is systematically different, which could bias effort estimates based on observed test scores.

Table 6: LATE of Middle School Policy on Retake Probability

	All Students (1)	Student with $DLP_i > 0$ (2)
τ_S	0.030***	0.034***
SE	0.005	0.009
Bandwidth (h)	0.43	0.41
Bias (L - R)	0.001 – 0	-0.001 – 0
Obs. (L - R)	23,974 – 55,801	10,128 – 30,844

Notes: The table reports the LATE of being eligible to attend an elite school on retake probability. Standard errors (SE) are robust.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.2 Effort and Test Scores

This section begins by presenting graphical evidence that eligibility for elite schools may incentivize students to increase their effort on the COMIPEMS entrance exam. First, I document a discontinuity in test scores at 7.0 middle school GPA eligibility threshold among students retaking the exam in 2016, which is not present among first-time takers in 2015. The remainder of the section introduces a series of RD designs. I begin estimating the canonical RD desing that assumes no selection. The framework is then extended to account for endogenous retake decisions by implementing models that adjust for sample selection, as well as approaches that relax the continuity assumption and rely on observable or unobservable characteristics.

4.2.1 Initial Evidence of Effort

In line with the limited time available for students to adjust their efforts, graphical inspection for students who took the COMIPEMS exam only once in 2015 (Figure 7) shows no visible discontinuity in test scores at the 7.0 GPA threshold. The same absence of a discontinuity is observed when examining the first exam attempt of students who eventually retook the exam, both for the full set of retakers (Figure 8a) and for the restricted retaker sample (Figure 9a). In contrast, in the following year, there is a discontinuity around the GPA seven threshold in middle school, indicating that students who retake the exam and have the opportunity to attend an elite school generally outperform those who lack this possibility in the neighborhood of the policy cutoff point (Figure 8b and Figure 9b). The graphical evidence suggests that eligibility for elite schools appears to be driving increases in test scores.

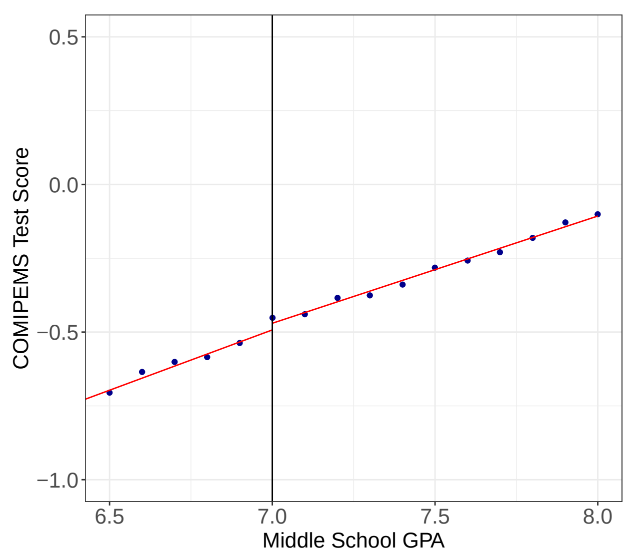
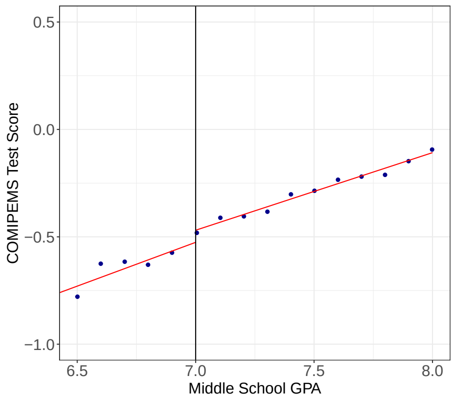
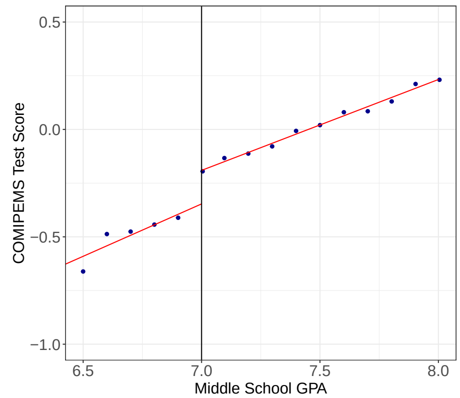


Figure 7: Discontinuity on Test Scores
First-time Takers, 2015

Notes: This figure displays the results from local polynomial regressions using the `rdplot` command. The horizontal axis represents students' middle school GPA, and the vertical axis shows normalized COMIPEMS entrance exam scores.



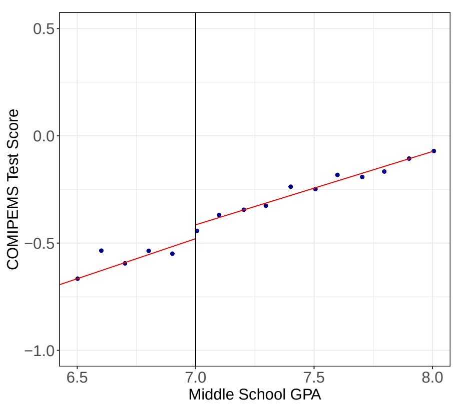
(a) 2015



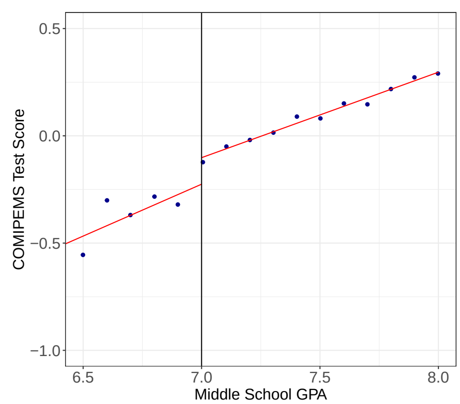
(b) 2016

Figure 8: Discontinuity on Test Scores
All Retakers

Notes: This figure displays the results from local polynomial regressions using the `rdplot` command. The horizontal axis represents students' middle school GPA, and the vertical axis shows normalized COMIPEMS entrance exam scores. Panel (a) includes the all retakers in 2015, while Panel (b) show the results for the all retakers in 2016.



(a) 2015



(b) 2016

Figure 9: Discontinuity on Test Scores on the Restricted Retakers Sample

Notes: This figure displays the results from local polynomial regressions using the `rdplot` command. The horizontal axis represents students' middle school GPA, and the vertical axis shows normalized COMIPEMS entrance exam scores. Panel (a) includes the restricted sample for retakers in 2015, while Panel (b) show the results for the same sample in 2016.

4.2.2 Canonical RD design

Table 7 shows the results, for the restricted sample, of the canonical RD design. In line with the limited time for students to adjust effort in the entrance exam, since they are unaware of their middle school GPA, for year 2015 there is no significant effect for the GPA policy.

In contrast, in 2016, were students are fully aware of their GPA, the estimates indicate a positive and significant effect of the policy on test scores. Students just above the 7.0 threshold scored between 0.23 and 0.25 standard deviations higher than those just below, in the standard and donut estimates. These results suggest that the policy may have led to increased academic effort. However, despite focusing on a subpopulation of students with clear stakes in accessing elite schools, these estimates may still be subject to bias due to sample selection, as test scores are only observed for students who chose to retake the exam.

Table 7: LATE on COMIPEMS Scores

	RDD 2015		RDD 2016	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
τ_O	0.077	0.095	0.229**	0.245**
SE	0.101	0.116	0.104	0.119
Bandwidth (h)	0.41	0.38	0.41	0.39
Bias (L - R)	0.013 – 0.003	0.016 – -0.011	0.058 – 0.003	0.052 – 0.003
Obs. (L - R)	1,154 – 4,192	979 – 2,566	1,149 – 4,189	973 – 2,563

Notes: This table reports the canonical RD design results. The estimates are computed for a restricted subpopulation that excludes multi-year retakers, students affiliated with the adult education system (INEA), out-of-state applicants, students who did not rank an elite school as their preferred option, and those unmatched or matched to elite schools in their first attempt in both 2015 and 2016. Donut specifications exclude students with a 7.0 middle school GPA. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.2.3 Difference and Lag Specification

The difference and lag specifications adjust for student heterogeneity by accounting for prior academic performance. This approach helps ensure that the observed effects of the policy are not confounded by pre-existing disparities in student ability. Table 8 shows that, under the difference specification, students just above the GPA threshold scored 0.10 and 0.09 standard deviations higher than those just below, in the standard and donut estimates, respectively. The lag specification yields similar results, with estimated impacts of 0.12 and 0.09 standard deviations for the standard and donut designs, respectively. Taken together, these findings reinforce the conclusion that the GPA policy led to a meaningful increase in academic performance among students who remained eligible to access elite institutions.

Table 8: LATE on COMIPEMS Scores
(2016, differences and lag)

	Diff		lag	
	Diff	Diff Donut	Lag	Lag Donut
	(1)	(2)	(3)	(4)
τ_O	0.102**	0.091*	0.115***	0.086*
SE	0.045	0.051	0.044	0.05
Bandwidth (h)	0.38	0.38	0.38	0.38
Bias (L - R)	0.032 – 0	0.035 – 0.016	0.035 – 0	0.037 – 0.014
Obs. (L -R)	973 – 3,339	973 – 2,563	973 – 3,339	973 – 2,563

Notes: This table reports the estimates from a regression discontinuity design using two alternative specifications. In the difference specification, the outcome variable is the difference in test scores. In the lag specification, the outcome is the 2016 test score, and the lagged (2015) test score is included as a covariate. The estimates are computed for a restricted subpopulation that excludes multi-year retakers, students affiliated with the adult education system (INEA), out-of-state applicants, students who did not rank an elite school as their preferred option, and those unmatched or matched to elite schools in their first attempt in both 2015 and 2016. Donut specifications exclude observations with a 7.0 middle school GPA. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

While the difference and lag specifications provide estimates of gains in test scores by adjusting for baseline performance, the earlier results reported in Table 7 do not account for prior achievement and instead capture level differences in test scores at the threshold. These level-based estimates, which range from 0.23 to 0.25 standard deviations in 2016, are larger than the gain-based estimates of 0.09 to 0.12 standard deviations reported in Table 8. This difference in magnitude reflects that the level specifications conflate the policy effect with potential differences in baseline ability or unobserved heterogeneity at the cutoff. In contrast, the gain estimates provide a cleaner measure of the policy-induced improvement in performance, isolating the extent to which students increased their academic effort as a response to the GPA threshold.

4.2.4 Sample Selection

Table 9 presents results addressing selection bias (Dong, 2019). I decompose the effect into extensive and intensive margins. The extensive margin captures the policy's impact on the decision to retake, and the intensive margin measures test-score effects among retakers. For the restricted sample, the extensive margin estimate is 2.6 percentage points, indicating that the GPA policy increases the probability of retaking. The intensive margin estimate is 0.23 sd. Under the monotonicity assumption, the corresponding sharp bounds range from a lower bound close to zero to an upper bound of 0.366.

Table 9: Extensive and Intensive Margins at the Cutoff

	Standard (1)	Donut (2)
$\Pr(S_0 = 1 \mid R = r_0)$	0.104***	0.100***
	(0.012)	(0.013)
$\Pr(S_1 = 1 \mid R = r_0)$	0.129***	0.127***
	(0.003)	(0.004)
Extensive margin (τ_S):	0.026	0.027
	(0.018)	(0.019)
$E[Y_0 \mid S_0 = 1, R = r_0]$	−0.413***	−0.410***
	(0.098)	(0.102)
$E[Y_1 \mid S_1 = 1, R = r_0]$	−0.181***	−0.174***
	(0.021)	(0.024)
Intensive margin (τ_O):	**0.232**	**0.236**
	(0.131)	(0.137)
Intensive Margin Bounds		
Lower bound:	0.004	0.005
Upper bound:	0.366	0.370

Notes: The table shows RD design estimations of the extensive margin, the increase in the probability of retake, and intensive margin, the increase in test scores. The intensive margin point estimate assumes no sample selection. Intensive margin bounds are built assuming monotonic selection. The donut specification exclude students with a 7.0 middle school GPA. Standard errors (SE) in parentheses are calculated using bootstrap resampling. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.2.5 Controlling for observables

The admission process questionnaire contains a wide range of questions that encompass student behavior, sociodemographic factors, and family material possessions that can be used as covariates in the regression discontinuity. Given that none of the variables display a discontinuity at the threshold of seven, I take an agnostic view and include all of them to fit a linear (Athey et al., 2019). I assume unconfoundedness (Rosenbaum and Rubin, 1983) which posits that once I control for the covariates, the treatment can be considered as good as randomly assigned, thereby enabling an unbiased estimate of, in this case, the LATE. The confidence intervals are obtained via bootstrap by resampling the data with replacement to estimate the distribution of the LATE. The findings indicate a median increase of 0.10 sd (0.12 for donut estimates) in the test scores of students to the right of the middle school GPA threshold, with a confidence interval ranging from 0.05 to 0.13 sd (0.7 to 0.14 for donut estimates) at a 95% confidence level.

4.2.6 Controlling for unobservables

Table 10 presents the results of the regression discontinuity analysis controlling for unobservable student characteristics by including as a covariate the proportion of elite schools to which students apply. The LATE coefficient corresponding to the 2015 estimates lacks statistical significance, which is consistent with prior findings. In contrast, the 2016 estimates remain significant. However, incorporating the variable that accounts for the proportion of elite schools to which students applied mitigates the impact of potential admission to elite schools from 0.229 and 0.245 for level and donut estimates (Table 7) to 0.179 and 0.155, respectively. This suggests that unobserved student characteristics may affect 2016 test scores, and including the proportion of elite schools in the students ROLs may correct otherwise biased estimates.[18]

Table 10: CATE on COMIPEMS Scores
(controlling for unobservables)

	RDD 2015		RDD 2016	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
τ_O	0.069	0.084	0.179***	0.155**
SE	0.063	0.074	0.069	0.078
Bandwidth	0.56	0.58	0.6	0.61
Bias	0.018 – 0.001	0.018 – -0.001	0.054 – 0.001	0.053 – 0.014
Obs.	1,154 – 4,192	979 – 2,566	1,149 – 4,189	973 – 2,563

Notes: This table reports the regression discontinuity results controlling for the proportion of elite schools listed in the rank-order. Estimates are based on the restricted subpopulation that excludes multi-year retakers, students affiliated with the adult education system (INEA), out-of-state applicants, students who did not rank an elite school as their preferred option, and those unmatched or matched to elite schools in their first attempt in both 2015 and 2016. Donut specifications exclude students with a 7.0 middle school GPA. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5 Validity

This section examines, for the restricted sample, the validity of the regression discontinuity design by testing for discontinuities in predetermined covariates, the lack of effects at placebo middle school GPAs, continuity, of the running variable, and the null impact among students who did not express interest in elite schools. While evidence of sample selection motivates the need for methods that address it, establishing the credibility of the regression discontinuity design still requires standard validity checks.

[18] Including the proportion of elite schools students listed in 2015 as a covariate decreases the LATE coefficient on the retaking probability by roughly half, from 0.030 to 0.017. Including covariates, such as mother and father educational level, student highest desired educational level, university the student desires to attend (e.g., UNAM, IPN, private, etc.), and fallback options in the event of rejection from a preferred institution, does not yield an impact in the LATE coefficient.

5.1 Null treatment effect on predetermined covariates

One of the key identifying assumptions in regression discontinuity designs is that there are no discontinuous jumps in potential outcomes or predetermined covariates at the cutoff, aside from the discontinuity induced by treatment assignment. This can be verified by checking if the covariates are balanced at the cutoff. The COMIPEMS survey provides a rich questionnaire on the characteristics of the students. The questionnaire addresses students' behavior, perceived socioeconomic status, family background, scholarships, and good and services they posses. Using the restricted sample, I inspect whether the covariates are balanced using a visual inspection and whether there is a jump using and RD design.

5.1.1 Students Behavior

The questionnaire assesses student behaviors related to establishing priorities, priorities order, deadlines, and anticipating important activities. Assessing these behaviors provides valuable insight into students time management and planning skills, which are key factors in maintaining consistent academic progress and achieving long-term educational goals. Students response in a discrete scale from one to four, one being the lowest value were student do not show the behavior often, and four showing it almost all the time. The visual inspection from Figure 10 shows no indication of discontinuities in student behavior, suggesting that these covariates are balanced at the middle school cutoff.

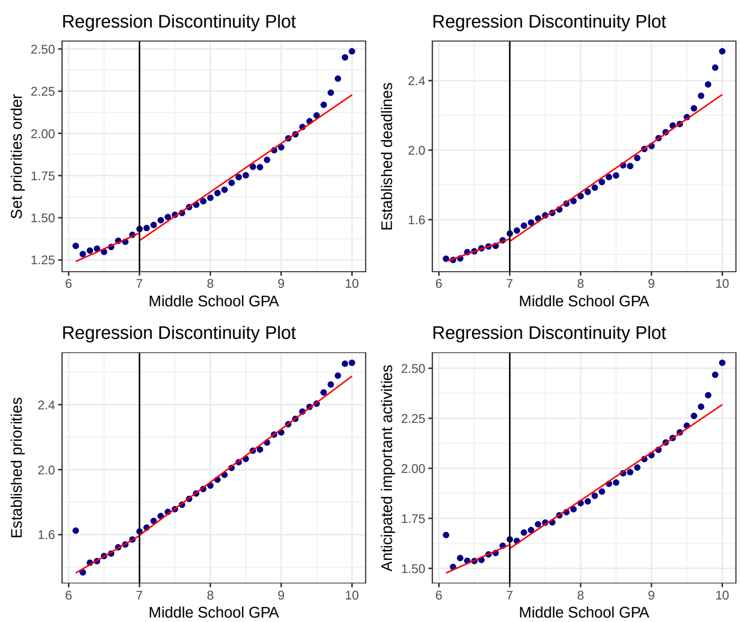


Figure 10: Covariates on students' behavior (2015)

Note: This figure displays the results from local polynomial regressions using the `rdplot` commnad for questionnaire based measures.

5.1.2 Sociodemographic and Scholarship

The questionnaire surveys students about their socioeconomic status, parental education, and whether they received a scholarship during their middle school studies. These factors contribute to explain educational outcomes. Perceived socioeconomic status shapes students' confidence, aspirations, and access to resources, often influencing students' academic performance. Parental education provides important foundations, as parents with higher education tend to offer more academic support and set higher expectations for their children's success. Mother's education, in particular, is often associated with early cognitive development and school readiness. Scholarships, whether for academic merit, financial need, or sports, help alleviate financial pressures and provide incentives for academic improvement. The visual inspection from Figure 11 reveals no apparent discontinuities in student behavior, suggesting that there are no systematic differences between students just above and below the threshold.

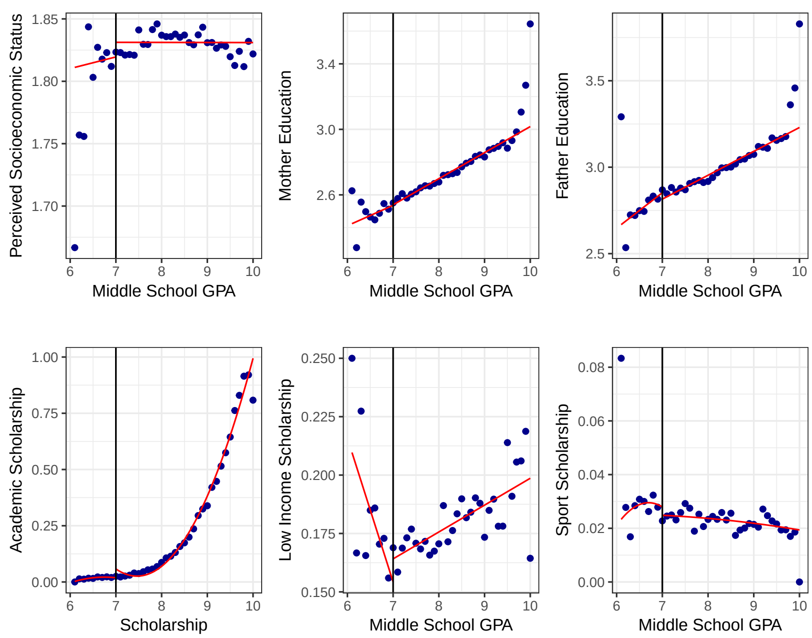


Figure 11: Sociodemographic and Scholarship Covariates (2015)

Notes: This figure shows local polynomial regressions using the `rdplot` command for questionnaire-based measures.

5.1.3 Good and Services

Goods and services such as the number of TVs, PCs, cars, bathrooms, annual vacations, tablets, landlines, washing machines, and other household items are not directly linked to educational performance, as these items themselves do not inherently affect a student's ability to learn or succeed academically. However, they can serve as indicators of socioeconomic status, which is often closely tied to educational outcomes. Higher access to such goods and services generally reflects a household's greater financial stability, access to resources, and overall living conditions. Families with better socioeconomic standing are more likely to provide supportive environments for learning, access to educational materials, and opportunities such as extracurricular activities, tutoring, or better schooling options. Therefore, while these goods and services do not directly influence education, they highlight the advantages that come with socioeconomic status, which can significantly impact a student's educational success.

The visual inspection of Figure 12 shows no clear discontinuities in the distribution of goods and services, indicating that there are no systematic differences in access to these items between students just above and below the threshold. This suggests that the availability of goods and services, such as TVs, PCs, cars, and other household items, remains consistent across groups.

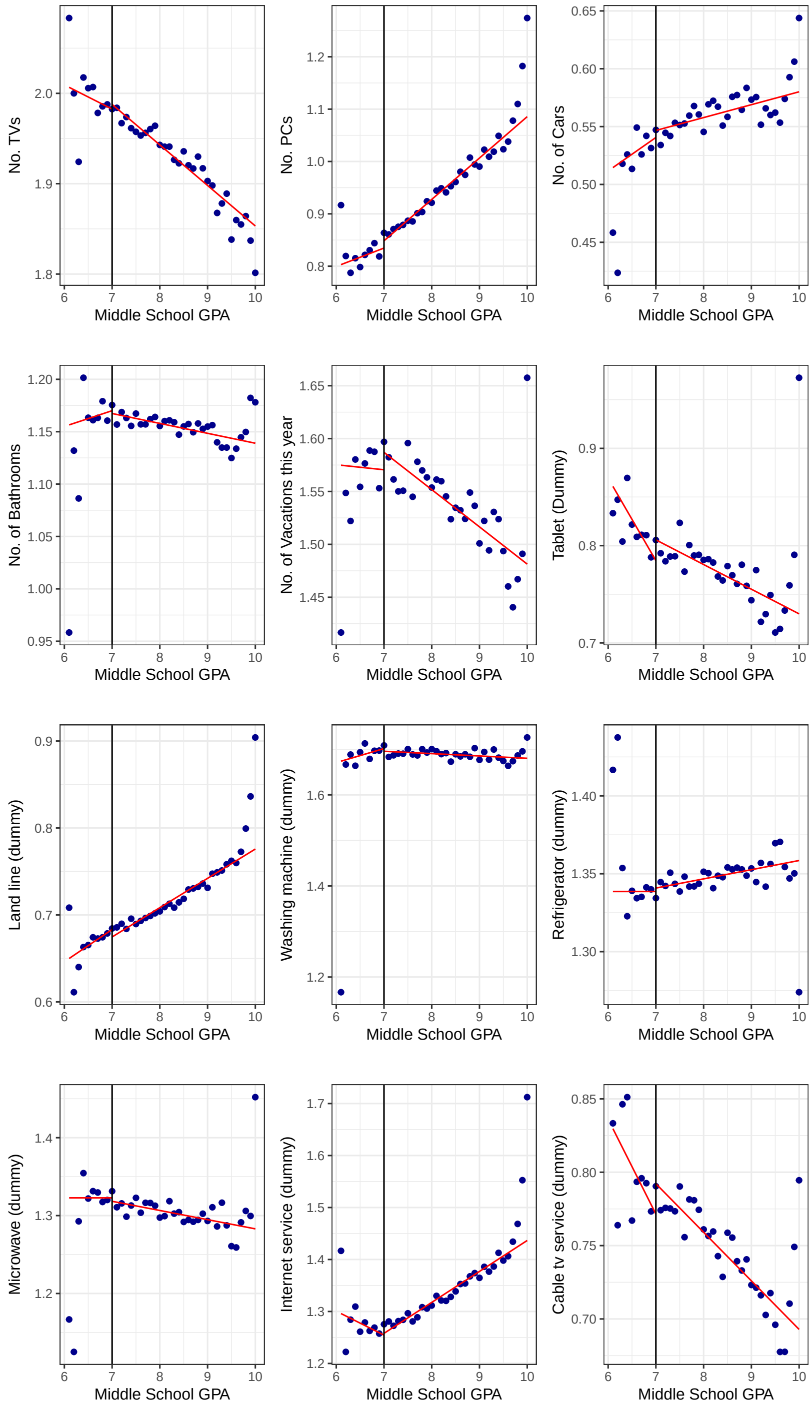


Figure 12: Covariates on goods and services (2015)

Notes: This figure shows local polynomial regressions using the `rdplot` command for questionnaire-based measures related to household goods and access to services.

5.1.4 RD design on Covariates

Table 11 reports RD estimates for a broad set of predetermined covariates at the 7.0 middle school GPA cutoff. The analysis covers three categories: student behavior, sociodemographic factors and scholarships, and goods and services. In the student behavior category, variables such as priority order, adherence to deadlines, ability to set priorities, and anticipation yield not significant coefficients. In the sociodemographic and scholarship category, most covariates, including perceived socioeconomic status, mother's education, and scholarship receipt, show no evidence of discontinuity, whereas father's education exhibits a statistically significant effect. In the goods and services category, most items, such as the number of televisions and vacations, are balanced, but PC ownership shows a statistically significant difference. The absence of systematic differences across covariates supports the validity of the design. The limited statistically significant results observed can be attributed to random variation, considering the extensive number of tests performed. When the Bonferroni p-value correction is applied, none of the coefficients maintain statistical significance.

Table 11: Regression Discontinuity Estimates on Covariates (2015)

Variable	τ
Student behavior	
Priority order	-0.034
Deadlines	0.007
Priorities	0.060
Anticipate plans	0.035
Sociodemographic and Scholarships	
Perceived socioeconomic status	-0.117
Mother education	0.148
Father education	0.392*
Academic scholarship	0.047
Income scholarship	0.085
Talent scholarship	0.015
Goods and Services	
# TVs	0.084
# PCs	0.267**
Car (dummy)	-0.024
# Bathrooms	0.129
# Vacations	0.105
Tablet (dummy)	0.167
Land line (dummy)	0.011
Washer (dummy)	0.062
Refrigerator (dummy)	-0.012
Microwave (dummy)	0.023
Internet (dummy)	0.122
Cable TV (dummy)	-0.004

Notes: Estimates are obtained from separate regression discontinuity (RDD) specifications using the restricted 2015 sample. Each row reports the estimated discontinuity at the GPA 7.0 threshold for the corresponding covariate, replacing the outcome variable in equation **??** with the respective 2015 COMIPEMS survey measure. Coefficients are reported without standard errors for brevity. Positive (negative) values indicate higher (lower) average values for students just above the cutoff relative to those just below. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.2 The treatment effect at artificial cutoffs values (placebo test)

Table 12 reports placebo RD estimates on the probability of retaking the COMIPEMS exam at GPA thresholds where the policy did not apply. The estimated LATEs are statistically significant at the 7.5 and 9.0 cutoffs, but not at 8.0, 8.5, or 9.5. The presence of significant effects at some placebo thresholds suggests that non-policy-related factors may influence retake behavior at certain points in the GPA distribution. This pattern warrants a cautious interpretation of the estimated discontinuity at 7.0, as it could partly reflect idiosyncratic variation unrelated to the policy. Further inspection of the distribution of retakers is therefore necessary to determine whether the observed jump at 7.0 reflects such idiosyncratic factors in addition to, or instead of, the policy effect. This analysis is performed in the next subsection 5.3.

Table 12: Falsifying Test on Probability of Retaking

	RDD				
	7.5	8.0	8.5	9.0	9.5
	(1)	(2)	(3)	(4)	(5)
τ	0.015**	0.001	0.007	0.014**	0.005
SE	0.006	0.005	0.005	0.006	0.006
Bandwidth	0.25	0.34	0.43	0.34	0.41
Bias	0.005-0	-0.002-0	0.001-0	0.002-0	0.003-0.001
Obs.	23,995-36,528	35,683-44,645	42,769-44,572	25,353-29,410	27,633-20,573

Notes: This table reports regression discontinuity estimates of the probability of retaking the COMIPEMS exam at different GPA thresholds. Each column corresponds to a different cutoff: 7.5, 8, 8.5, 9, and 9.5. The reported estimates correspond to the local average treatment effect (LATE) from a canonical RDD model using a linear specification. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 13 presents the corresponding placebo tests for COMIPEMS test scores. Across all placebo thresholds, the estimated coefficients are negative, small in magnitude, and statistically insignificant. The absence of positive or significant effects at any non-policy cutoff strengthens the credibility of the main findings, as the significant improvement in test scores at the 7.0 threshold is not mirrored elsewhere in the GPA distribution.

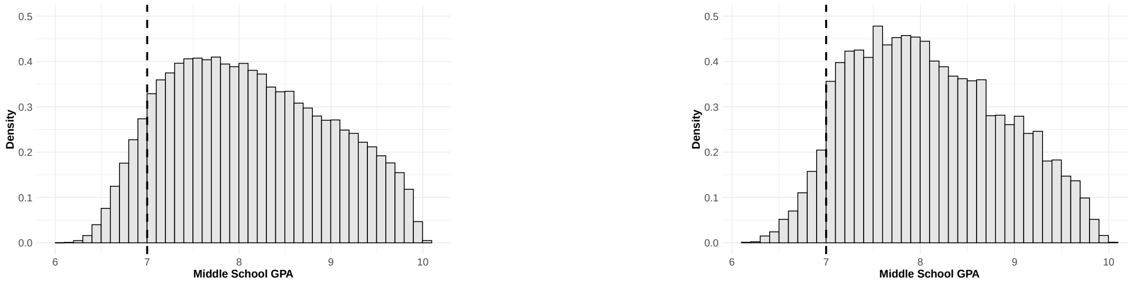
Table 13: Placebo Test Score RDD

	RDD				
	7.5	8.0	8.5	9.0	9.5
	(1)	(2)	(3)	(4)	(5)
τ	-0.027	-0.035	-0.075	-0.062	-0.074
SE	0.066	0.063	0.073	0.073	0.073
Bandwidth	0.41	0.37	0.33	0.34	0.42
Bias	0.009–-0.001	-0.016–0.002	0.001–0.005	0.022–0.002	-0.018–-0.009
Obs.	3,413–4,534	2,654–3,359	2,300–2,702	1,798–2,111	1,929–1,097

Notes: This table reports regression discontinuity estimates of COMIPEMS test scores at different GPA thresholds. Each column corresponds to a different cutoff: 7.5, 8, 8.5, 9, and 9.5. The reported estimates correspond to the local average treatment effect (LATE) from a canonical RDD model using a linear specification. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

5.3 Continuity

Continuity at the 7.0 middle school GPA provides credibility against self-selection, it guards against sorting or manipulation and supports comparability between students above and below the GPA policy. By contrast, a density jump, signals sorting at the margin, so any outcome discontinuity cannot be attributed solely to the policy. Establishing discontinuities in the distribution under the current setup is a challenging task for three main reasons. Firstly, the running variable is coarse, as GPA grading occurs at 0.1 intervals, thereby lacking true continuity. Consequently, it is not possible to approximate the densities to the right $f^+(r)$ and left $f^-(r)$ of a cutoff point (r) using standard kernel methods, and discrete support of the running variable can introduce artificial jumps between adjacent GPA values. Secondly, the number of students above and below arbitrary middle school GPAs can display abrupt jumps that reflect grading conventions rather than policy effects. Such discontinuities may arise simply because teachers tend to cluster grades around certain values (ter Meulen, 2023; Bowden et al., 2023) or bunching students at a passing grade (Diamond and Persson, 2016). Thirdly, if discontinuities are already present in the full sample due to institutional conventions or any unrelated policy intervention, observed discontinuities among retakers could potentially reflect pre-existing patterns rather than strategic manipulation.



(a) 2015 First-time Takers

(b) 2016 Restricted Retakers Subsample

Figure 13: Running Variable Densities

Notes: This figure shows histogram of the running variable using 0.1 wide bins. Panel (a) shows the all first-time takers in 2015; Panel (b) show the restricted sample of retakers.

To inspect discontinuities around the threshold I explore the histograms for first-time takers and retakers (Figure 13) using 0.1 wide bins that exactly match the coarseness at which GPA is recorded. The distribution for first-time takers (Figure 13a) is smoother overall. By contrast, the retaker distribution (Figure 13b) is more jagged and exhibits a substantial jump at the 7.0 cutoff. Notwithstanding its limitations in addressing the issues identified, to test whether the distributions exhibit discontinuities I implement McCrary (2008) test, the standard diagnostic for manipulation in RD design applications, at $r \in \{7.0, 7.1, 7.2, \ldots, 9.4, 9.5\}$ using a bandwidth of 0.5. To ensure that the reported discontinuities are not spurious, given the coarse 0.1 increments of the running variable, I impose a stricter criterion and only report significant results at the 1% level. Table 14 reports the results. For first-time takers, statistically significant discontinuities are found at GPA cutoffs of 7.0, 8.0, 8.3, 8.5, and 9.0. Except for the 8.3 cutoff point, all align with grading values that are attractive focal points for teachers, such as whole numbers or half-point marks, consistent with institutional grading conventions rather than abrupt changes in the underlying student performance. For retakers, the reduced number of observations lowers statistical power. Nonetheless, a discontinuity persists at the 7.0 threshold, while other focal grading values display no observable discontinuities.

Table 14: McCrary Density Test by Cutoff

GPA	First-time takers	Retakers
7.0	$\star$	$\star$
8.0	$\star$	
8.3	$\star$	
8.5	$\star$	
8.7		$\star$
9.0	$\star$	

Notes: This table reports results from the McCrary (2008) density test at various middle school GPA thresholds. $\star$ denotes significance at the 1% level.

Given the Mccrary test results, a remaining concern is that the discontinuities reflect baseline composition differences and not retakers sizable effects in the running variable. To address this, I implement Fitzgerald (2025) test equivalence procedure, which evaluates whether any discontinuity is meaningfully large. Let $f^+(r_0)$ and $f^-(r_0)$ denote the right and left limits of the running variable density at cutoff r_0, and define the right to left ratio $\rho(r_0) = f^+(r_0)/f^-(r_0)$ with $\theta = \log \rho(r_0)$. The test constructs an equivalence confidence interval (ECI) for $\hat{\theta}$ from the data, and compare it with a researcher chosen ratio $\varepsilon = f^+(r_0)/f^-(r_0) > 1$.In the absence of specific data regarding the anticipated magnitude of the discontinuity (information about ε). I report the upper and lower ECI bounds.

Figure 14 plots the Fitzgerald equivalence bound by cutoff. From 7.0 to 8.0 I display the upper bound ratio consistent with the maximum upward jump admissible in the density. From 8.1 onward I plot the lower bound consistent with the maximum admissible drop. The procedure is symmetric. For instance, a right to left density ratio of 2 is equivalent to a 100% upward jump, whereas a right to left density ratio of 1/2 is equivalent to a 50% downward jump. At $r_0 = 7.0$, the retaker sample yields a maximum ratio of 1.85, which implies a maximum jump of 85%. By contrast, for first-time takers at the same cutoff the maximum jump is calculated at 1.10. Therefore, density jumps exceeding 10% have a small chance to occur. The bounds do not provide a formal test but provide reassurance that the jump is indeed large enough for the selection approach to be well grounded.

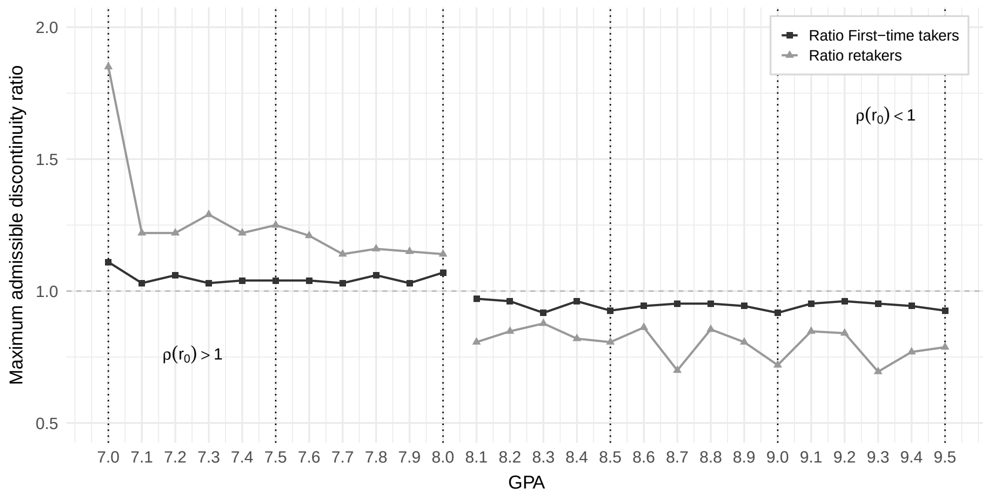
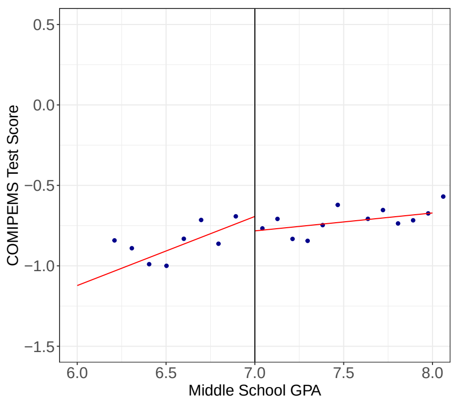


Figure 14: Fitzgerald Bounds

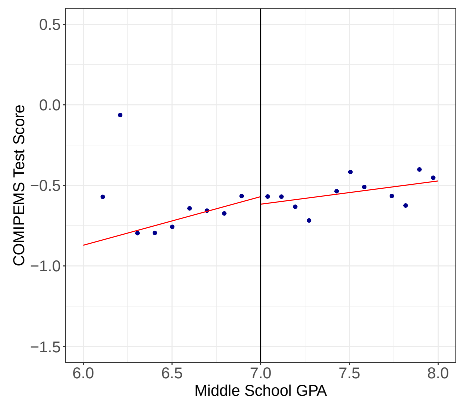
Note: This figure plots Fitzgerald equivalence confidence intervals (ECIs) at the 5% level by GPA cutoff. For values between 7.0 and 8.0, the figure plots the right-to-left upper bound density ratio, where values greater than one correspond to upward jumps. For cutoffs above 8.0, the plot provides left-to-right ratio, where values below one correspond to downward jumps.

5.4 Students not interested in Elite Schools

A discontinuity at the 7.0 threshold among students who did not express interest in elite schools would indicate that forces other than elite-school aspirations contribute to the observed jump in COMIPEMS test scores. Figure 15 provides no indication of such a discontinuity for students not interested in elite schools, and this observation is supported by Table 15, which presents estimation results that reveal no significant effects for years 2015 and 2016 in both the level and the donut estimates. The absence of a discontinuity can be attributed to their indifference towards the high stakes associated with elite educational pathways. Thus, the academic performance of these students reflects a steady and consistent level of effort that does not change significantly as a result of approaching the GPA threshold.



(a) 2015



(b) 2016

Figure 15: Regression Discontinuity on Test Scores for Students not Interested in Elite Schools

Table 15: LATE on COMIPEMS Scores
(Students not Interested in Elite Schools)

	RDD 2015		RDD 2016	
	(1) Level	(2) Donut	(3) Level	(4) Donut
Coefficient	-0.135	0.081	-0.092	0.085
SE	0.14	0.182	0.134	0.183
Bandwidth	0.37	0.32	0.39	0.33
Bias	-0.053 – -0.007	-0.025 – -0.124	-0.061 – -0.011	-0.038 – -0.128
Effective obs.	296 – 409	296 – 306	373 – 432	373 – 321

Notes: This table reports RD design results. Donut specifications exclude students with a 7.0 middle school GPA. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

6 Financial Investment and Socioemotional Indicators

The COMIPEMS survey provides information about how students prepare for the exam, asking them if they took courses at their school, with a private institution, or with a tutor. The responses are recorded as dichotomous variables, coded one if the student engages in the preparatory format, and zero otherwise. In 2015, 33.0% of students reported taking courses at their school, 32.2% enrolled in courses offered by private institutions, and 17.8% received support from a private tutor. To assess whether eligibility at the 7.0 GPA cutoff affected these preparation choices, I estimate the canonical RD design using a linear probability model.

Table 16 provides the results for the linear probability model. There is no difference for years 2015 in courses taken in school, private institutions or with a private tutor between students with the opportunity to access elite institutions from those without it. This is consistent with students no being aware of their middle school GPA. At the discontinuity students that are not eligible to attend elite schools are 15.1% percentage points more likely to take courses at their schools, while students that are eligible to attend elite schools are 14.4% more likely to take courses in private institutions. Courses offered by students' schools are generally free or less pricey than private courses.[19] Although both groups of students continue to prepare for the exam, their methods differ. When afforded eligibility to attend an elite school, families devote more financial resources to exam preparation than families of otherwise similar students who lack eligibility. The RD design analysis of classes with private tutors yields no significant results, and the donut approach corroborates the baseline results.[20]

Table 16: LATE on the Probability of Taking Courses

	RDD 2015		RDD 2016	
	Level	Donut	Level	Donut
	(1)	(2)	(3)	(4)
in school	-0.058	-0.048	-0.151***	-0.097*
	(0.066)	(0.074)	(0.058)	(0.055)
in private institution	0.082	0.086	0.144**	0.135*
	(0.065)	(0.074)	(0.070)	(0.079)
private tutor	0.012	-0.011	-0.022	-0.012
	(0.055)	(0.062)	(0.060)	(0.068)

Notes: This table reports canonical RD LATE estimates at the 7.0 middle school GPA cutoff for the probability of taking courses in school, at a private institution, or with a private tutor for the high school entrance exam. Donut specifications exclude observations with GPA exactly 7.0.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

In addition to academic investment, I also examine a set of socio-emotional variables collected in the COMIPEMS survey about whether the students face anxiety, aggression, depression or attention problems. The variables are recorded dichotomously, assigning a value of one to students who demonstrate the specified behavior and zero to those who do not. Table 17 shows the RD design estimates for the linear probability model. I do not find any significant results being eligible or non eligible to elite school in students having anxiety,

[19] In 2025, private COMIPEMS preparation courses cost between \$155–\$565 USD. By contrast, a free public alternative Puntos de Innovación, Libertad, Arte, Educación y Saberes (PILARES, by its acronym in Spanish) in Mexico City, now offers no, fee preparation. However, these municipal centers began operating in 2019 and thus were not available during our 2015–2016 study period. Throughout the duration of our study, certain schools provided COMIPEMS courses either free of charge or for a small nominal fee. Dollar amounts are approximate and converted using the Banxico FIX of 18.784 MXN/USD (August 19, 2025).

[20] The robustness of the RD design is confirmed across specifications employing differences, lags, and controls unobservables. The random forest approach faces collinearity issues due to the dummy nature of the variables. Bounds are calculated and are too wide to be informative. See Appendices A2 and A3.

aggression, depression or attention problems. This pattern contrasts with evidence from other highly competitive admission settings that document elevated stress and worse mental-health symptoms around entrance examinations or complex admissions regimes (Chen et al., 2022; Xu and Lee, 2023).

Several features of this setting can explain why our results differ from studies documenting adverse mental health effects around highly competitive entrance exams. First, I study high school eligibility, not college admission, were college admission are consider higher stake. Second, the COMIPEMS questionnaire is a general survey rather than a purpose built mental health instrument; these binary indicators are coarse and may miss variation that specialized scales capture. Third, the RD design identifies a local effect at the 7.0 GPA threshold, which lies in the lower segment of the GPA distribution, stress responses among high achievers students may provide different results.

Table 17: LATE on the Probability of Exhibiting Socioemotional Behaviors

	2015		2016	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
Anxiety	0.024	0.041	-0.017	0.005
	(0.035)	(0.039)	(0.039)	(0.044)
Aggression	0.016	0.003	0.017	0.018
	(0.022)	(0.023)	(0.019)	(0.020)
Depression	-0.008	0.001	-0.008	-0.005
	(0.023)	(0.025)	(0.023)	(0.026)
Attention	0.005	0.012	-0.011	-0.018
	(0.032)	(0.035)	(0.034)	(0.036)

Notes: This table reports canonical RD LATE estimates at the 7.0 middle school GPA cutoff for socio-emotional indicators (anxiety, aggression, depression, and attention). Donut specifications exclude observations with GPA exactly 7.0.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

7 Conclusion

This paper examines how eligibility for elite high schools in Mexico's centralized admission system shapes student effort and retake behavior. Exploiting the sharp GPA threshold of 7.0, I show that students just above the cutoff are more likely to retake the COMIPEMS exam and, conditional on retaking, achieve higher test scores than their counterparts below the threshold. These results provide direct evidence that access to elite educational opportunities motivates additional effort. Beyond effort, the findings reveal important implications for family educational investment. Families of eligible students devote greater financial resources to private preparatory courses, while ineligible students rely more heavily on free school-based preparation. In addition, I do not find any evidence of effects on self-reported socioemotional outcomes such as anxiety or depression. These null effects should be interpreted cautiously, they may suggest that socioemotional responses may be more muted for students at the lower end of the GPA distribution or that the survey instruments were insufficiently sensitive to capture mental health consequences.

From a policy perspective, the results demonstrate that the GPA requirement functions as a lever that shapes student incentives. Conditioning access to elite schools on prior grades increases preparation among eligible students while discouraging those just below the cutoff. This mechanism illustrates how admissions rules not only allocate scarce seats but also direct the intensity and form of effort students and families undertake. Recognizing this dual role is critical: poorly designed thresholds risk generating wasted preparation, whereas well-calibrated policies can channel effort more productively and equitably.

The study shows a direct link between students' ROL and their exam effort, providing evidence of how admissions incentives shape behavior. This contribution is particularly relevant for studies on centralized matching mechanisms, where exam performance determines access to selective schools but the role of student effort is often unobserved (Agarwal and Somaini, 2020). To the best of my knowledge, only a few studies explicitly incorporate effort into the student decision-making process (e.g., Arslan, 2022), suggesting a promising direction for future research.

References

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2014). The elite illusion: Achievement effects at boston and new york exam schools. *Econometrica*, 82(1):137–196.

Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., and Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, 110(5):1502–1539.

Agarwal, N. and Somaini, P. (2020). Revealed preference analysis of school choice models. *Annual Review of Economics*, 12(1):471–501.

Ajayi, K. F. (2024). School choice and educational mobility: Lessons from secondary school applications in ghana. *Journal of Human Resources*, 59(4):1207–1243.

Akyol, P. and Krishna, K. (2017). Preferences, selection, and value added: A structural approach. *European Economic Review*, 91:89–117.

Almond, D., Doyle Jr, J. J., Kowalski, A. E., and Williams, H. (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. *The quarterly journal of economics*, 125(2):591–634.

Andrabi, T., Das, J., Ijaz Khwaja, A., and Zajonc, T. (2011). Do value-added estimates add value? accounting for learning dynamics. *American Economic Journal: Applied Economics*, 3(3):29–54.

Angrist, J. D. and Rokkanen, M. (2015). Wanna get away? regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association*, 110(512):1331–1344.

Arcidiacono, P., Aucejo, E. M., and Hotz, V. J. (2016). University differences in the graduation of minorities in stem fields: Evidence from california. *American Economic Review*, 106(3):525–562.

Arslan, H. A. (2022). An empirical analysis of college admissions with endogenous entrance exam scores. *Available at SSRN 3346459*.

Artemov, G., Che, Y.-K., and He, Y. (2017). Strategic 'mistakes': Implications for market design research. *NBER working paper*.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *Annals of Statistics*, 47(2).

Biró, P. (2007). Higher education admission in hungary by a score-limit algorithm. In *The 18th International Conference on Game Theory at Stony Brook University*.

Bjork, C. (2011). Imagining japan's 'relaxed education'curriculum: Continuity or change. *Reimagining Japanese education: Borders, transfers, circulations, and the comparative*, pages 147–169.

Bobba, M., Frisancho, V., and Pariguana, M. (2023). Perceived ability and school choices: Experimental evidence and scale-up effects.

Bowden, A. B., Rodriguez, V., and Weingarten, Z. (2023). The unintended consequences of academic leniency. edworkingpaper no. 23-836. *Annenberg Institute for School Reform at Brown University*.

Cabrera-Hernández, F., Padilla-Romo, M., and Peluffo, C. (2023). Full-time schools and educational trajectories: Evidence from high-stakes exams. *Economics of Education Review*, 96:102443.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Chen, F., Chen, J., Chen, B., Mofatteh, M., Wen, C., Wellington, J., Gong, D., Yang, H., Zeng, Z., Miao, X., et al. (2022). Mental health status of medical students during postgraduate entrance examination. *BMC psychiatry*, 22(1):829.

Chen, L. and Pereyra, J. S. (2019). Self-selection in school choice. *Games and Economic Behavior*, 117:59–81.

Chen, Y. and Kesten, O. (2017). Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, 125(1):99–139.

Dale, S. B. and Krueger, A. B. (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of human resources*, 49(2):323–358.

Dang, H.-A. and Rogers, F. H. (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *The World Bank Research Observer*, 23(2):161–200.

DellaVigna, S. and Pope, D. (2018). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

Denning, J. T., Eide, E. R., Mumford, K. J., Patterson, R. W., and Warnick, M. (2022). Why have college completion rates increased? *American Economic Journal: Applied Economics*, 14(3):1–29.

Diamond, R. and Persson, P. (2016). The long-term consequences of teacher discretion in grading of high-stakes tests. Technical report, National Bureau of Economic Research.

Dong, Y. (2019). Regression discontinuity designs with sample selection. *Journal of Business & Economic Statistics*, 37(1):171–186.

Dustan, A., De Janvry, A., and Sadoulet, E. (2017). Flourish or fail?: The risky reward of elite high school admission in mexico city. *Journal of Human Resources*, 52(3):756–799.

Ehrenberg, R. G. and Bognanno, M. L. (1990). Do tournaments have incentive effects? *Journal of political Economy*, 98(6):1307–1324.

Entrich, S. R. (2017). *Shadow education and social inequalities in Japan: Evolving patterns and conceptual implications*. Springer.

Estrada, R. (2017). The effect of the increasing demand for elite schools on stratification.

Estrada, R. and Gignoux, J. (2017). Benefits to elite schools and the expected returns to education: Evidence from mexico city. *European Economic Review*, 95:168–194.

Evans, D. and Yuan, F. (2019). Equivalent years of schooling: A metric to communicate learning gains in concrete terms. *World Bank Policy Research Working Paper*, (8752).

Fitzgerald, J. (2025). Manipulation tests in regression discontinuity design: The need for equivalence testing.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(3):447–456.

Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., and Xu, Y. (2019). Measuring success in education: The role of effort on the test itself. *American Economic Review: Insights*, 1(3):291–308.

Gneezy, U. and Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly journal of economics*, 115(3):791–810.

Goodman, J., Gurantz, O., and Smith, J. (2020). Take two! sat retaking and college enrollment gaps. *American Economic Journal: Economic Policy*, 12(2):115–158.

Guarino, C. M., Reckase, M. D., Stacy, B. W., and Wooldridge, J. M. (2015). Evaluating specification tests in the context of value-added estimation. *Journal of Research on Educational Effectiveness*, 8(1):35–59.

Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161.

Hiller, V. and Tercieux, O. (2014). Choix d'écoles en france. *Revue économique*, 65(3):619–656.

Horowitz, J. L. and Manski, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, pages 281–302.

Instituto Kepler (2023). Estadísticas del proceso de admisión ipn nivel superior. https://institutokepler.com.mx/estadisticas-del-proceso-de-admision-ipn-nivel-superior/. Accessed: 2023-09-30.

IPN (2022). Informe anual de actividades 2022. https://www.ipn.mx/assets/files/coplaneval/docs/dii/informe_anual_actividades/informe-anual-2022.pdf. PDF file.

Karadal, H., Celik, C., and Saygın, M. (2013). Corporate values on strategic planning process: A research about the universities in turkey. *Procedia-Social and Behavioral Sciences*, 99:762–770.

Kim, S. and Lee, J.-H. (2010). Private tutoring and demand for education in south korea. *Economic development and cultural change*, 58(2):259–296.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The quarterly journal of economics*, 131(3):1057–1111.

Kolesár, M. and Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8):2277–2304.

Krishna, K., Lychagin, S., and Frisancho, V. (2018). Retaking in high stakes exams: Is less more? *International Economic Review*, 59(2):449–477.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.

Liu, L. and Neilson, W. S. (2011). High scores but low skills. *Economics of Education Review*, 30(3):507–516.

Luflade, M. (2018). *Essays on the Empirics of School Choice*. PhD thesis, Duke University.

Machado, C. and Szerman, C. (2016). Centralized admission and the student-college match. Technical report, IZA Discussion Papers.

McCaffrey, D. F., Sass, T. R., Lockwood, J., and Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education finance and Policy*, 4(4):572–606.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714.

Ngo, D. and Dustan, A. (2024). Preferences, access, and the stem gender gap in centralized high school assignment. *American Economic Journal: Applied Economics*, 16(4):257–287.

Ortega Hesles, M. E. (2015). *School choice and educational opportunities: The upper-secondary student-assignment process in Mexico City*. PhD thesis.

Pariguana, M. and Ortega, M. (2022). *School Admissions, Mismatch, and Graduation*. PhD thesis, PhD thesis, University of Western Ontario.

Peng, S. and Ning, Y. (2021). Regression discontinuity design under self-selection. In *International Conference on Artificial Intelligence and Statistics*, pages 118–126. PMLR.

Pop-Eleches, C. and Urquiola, M. (2013). Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289–1324.

Qian, H., Walker, A., and Chen, S. (2024). The 'double-reduction'education policy in china: Three prevailing narratives. *Journal of Education Policy*, 39(4):602–621.

Rios, I., Larroucau, T., Parra, G., and Cominetti, R. (2021). Improving the chilean college admissions system. *Operations Research*, 69(4):1186–1205.

Rosenbaum, P. R. and Rubin, D. B. (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B (Methodological)*, 45(2):212–218.

Stinebrickner, R. and Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The BE Journal of Economic Analysis & Policy*, 8(1).

Teo, C.-P., Sethuraman, J., and Tan, W.-P. (2001). Gale-shapley stable marriage problem revisited: Strategic issues and applications. *Management Science*, 47(9):1252–1267.

ter Meulen, S. (2023). Long-term effects of grade retention. Technical report, CESifo Working Paper.

UNAM (2023). Series estadísticas. https://estadistica.unam.mx/index.php/series-estadisticas/. Accessed: 2023-09-30.

Wang, T. and Zhou, C. (2020). High school admission reform in china: a welfare analysis. *Review of Economic Design*, 24(3):215–269.

Xu, J. and Lee, S. G. (2023). Problems with complex college admissions policies and overloaded after-school private education on middle-and high-school students' mental health in south korea. *Journal of Global Health Reports*, 7:e2023049.

Yang, W., Sun, R., Wang, C., Chen, J., Zhang, C., Yu, J., and Liu, H. (2023). Epidemiology of depressive disorders among youth during gaokao to college in china: results from hunan normal university mental health survey. *BMC psychiatry*, 23(1):481.

Yatkin, E., Aral, N., Gunes, L. C., and Tosun, S. (2023). Mental well-being and test anxiety among students preparing for the university admission exam during the pandemic. *Frontiers in Psychology*, 14:1184788.

Zhang, Y. (2013). Does private tutoring improve students' national college entrance exam performance?—a case study from jinan, china. *Economics of Education Review*, 32:1–28.

A1 GPA Sensitivity at the End of Middle School

In the Mexican secondary school system, students typically take nine courses each year, and their final GPA is based on a series of bimonthly grades recorded throughout the three years of middle school. By the end of the third year, students accumulate more than a hundred grades, each contributing to the final GPA. As the academic year draws to a conclusion, most of the grades are already fixed, leaving students with limited room to influence their final GPA. For those whose GPA hovers around the critical 7.0 threshold for elite eligibility, assuring a 7.0 middle school GPA posses a high stakes situation.

The implementation of SIGED in 2013 (Sistema de Información y Gestión Educativa) that standardized the process by which grades are reported and recorded in public schools in Mexico, limit students option to rig the system. In this digital platform, teachers submit grades directly into an online system, reducing opportunities for informal grade negotiation after the official grading period closes. In the final two-month period, often the last opportunity to improve, students can still try to influence their results. However, achieving even a small change in the final GPA would require consistently high marks across multiple subjects. In practice, if the student does not change her commitment, this means that the student must persuade several teachers to grant higher scores. A back of the envelope calculation indicates that a student with a prior GPA of 6.9 would need to convince at least four teachers to award perfect scores of 10 in the final grading period in order to reach a 7.0 GPA, or alternatively persuade all nine teachers to grant just one additional point above their current performance of 6.9.

A2 LATE on Probability of Taking Courses

Table A1: LATE on the Probability of Taking Courses
(difference and lag)

	diff		lag	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
in school	-0.118	-0.058	-0.150***	-0.095*
	(0.093)	(0.092)	(0.058)	(0.055)
in private institution	0.063	0.004	0.128*	0.111
	(0.091)	(0.099)	(0.070)	(0.078)
private tutor	0.048	0.050	0.006	0.004
	(0.092)	(0.091)	(0.070)	(0.069)

Notes: This table reports RD design estimates at the 7.0 middle school GPA cutoff for the probability of taking courses in school, at a private institution, or with a private tutor in 2016. Each entry shows the robust bias-corrected local linear estimate (top line) with robust standard error in parentheses. The difference specification uses the change in course-taking between the current year and the lag (2016 − 2015). The lag specification uses the 2016 indicator as the outcome and includes the 2015 lag as a covariate. Donut specifications exclude students with a GPA exactly equal to 7.0. Bandwidths are MSE-optimal, and robust bias-corrected (RBC) inference is reported. Because the running variable is discrete, nearest-neighbor variance estimation is employed.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A2: LATE on the Probability of Taking Courses
(controlling for unobservables)

	RDD 2015		RDD 2016	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
in school	-0.026	0.008	-0.148***	-0.098
	(0.078)	(0.089)	(0.055)	(0.062)
in private institution	0.051	0.097	0.122*	0.106
	(0.077)	(0.089)	(0.068)	(0.090)
private tutor	-0.030	-0.073	-0.011	0.025
	(0.065)	(0.075)	(0.069)	(0.079)

Notes: This table reports RD estimates at the 7.0 middle school GPA cutoff for the probability of taking courses in school, at a private institution, or with a private tutor, controlling for the proportion of elite schools in students' rank-order lists. Donut specifications exclude observations with GPA exactly 7.0.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A3: Intensive Margin Estimates on the Probability of Taking Courses

	in school		in private institution		private tutor	
	standard	donut	standard	donut	standard	donut
Intesive Margin	-0.151	-0.097	0.144	0.135	-0.022	-0.012
bounds	[-0.61, 0.25]	[-0.52, 0.28]	[-0.42, 0.67]	[-0.47, 0.66]	[-0.55, 0.49]	[-0.58, 0.50]

Notes: The table shows the intensive margin estimate from RDD at the 7.0 middle school GPA cutoff and the corresponding bounds under monotonic selection.

A3 LATE on Probability of Exhibiting Socioemotional Behaviors

Table A4: LATE on the Probability of Exhibiting Socioemotional Behaviors (difference and lag)

	diff		lag	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
Anxiety	-0.047	-0.045	-0.025	-0.010
	(0.045)	(0.051)	(0.037)	(0.042)
Aggression	-0.001	0.015	0.015	0.018
	(0.029)	(0.030)	(0.019)	(0.020)
Depression	-0.013	-0.017	-0.008	-0.008
	(0.028)	(0.032)	(0.023)	(0.025)
Attention	-0.015	-0.030	-0.012	-0.020
	(0.041)	(0.045)	(0.032)	(0.035)

Notes: This table reports RD design estimates at the 7.0 middle school GPA cutoff for socio-emotional indicators (anxiety, aggression, depression, and attention) in 2016. Each entry shows the robust bias-corrected local linear estimate (top line) with robust standard error in parentheses. The difference specification uses the change in the indicator between the current year and the lag ($2016 - 2015$). The lag specification uses the 2016 indicator as the outcome and includes the 2015 lag as a covariate. Donut specifications exclude students with a GPA exactly equal to 7.0. Bandwidths are MSE-optimal, and robust bias-corrected (RBC) inference is reported. Because the running variable is discrete, nearest-neighbor variance estimation is employed.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A5: LATE on the Probability of Exhibiting Socioemotional Behaviors (controlling for unobservables)

	RDD 2015		RDD 2016	
	Standard	Donut	Standard	Donut
	(1)	(2)	(3)	(4)
anxiety	0.021	0.044	-0.018	0.017
	(0.040)	(0.047)	(0.045)	(0.052)
aggression	0.014	0.007	0.018	0.022
	(0.029)	(0.032)	(0.022)	(0.025)
depression	-0.016	-0.009	-0.018	-0.012
	(0.026)	(0.030)	(0.027)	(0.031)
attention	0.012	0.015	-0.000	-0.001
	(0.039)	(0.044)	(0.040)	(0.044)

Notes: This table reports RD estimates at the 7.0 middle school GPA cutoff for socio-emotional indicators (anxiety, aggression, depression, and attention) in 2015 and 2016, controlling for the proportion of elite schools in students' rank-order lists. Each entry shows the robust bias-corrected local linear estimate (top line) with robust standard error in parentheses. Donut specifications exclude observations with GPA exactly equal to 7.0.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Intensive Margin Estimates on the Probability of Exhibiting Socioemotional Behaviors

	anxiety		aggression	
	standard	donut	standard	donut
estimate	-0.017	0.005	0.017	0.018
bounds	[-0.45, 0.44]	[-0.42, 0.47]	[-0.33, 0.39]	[-0.34, 0.38]
	depression		attention	
	standard	donut	standard	donut
estimate	-0.008	-0.005	-0.011	-0.018
bounds	[-0.37, 0.36]	[-0.36, 0.35]	[-0.44, 0.40]	[-0.45, 0.41]

Notes: The table shows the intensive margin estimate from RDD at the 7.0 middle school GPA cutoff and the corresponding bounds under monotonic selection.