Title
Exploring Education Policies and their Socioeconomic Implications: Evidence and Insights from Juvenile Crime, College Admissions, and Gender Bias

Author
Reyes Hinrichsen, Tatiana

Publication Date
2023

Peer reviewed|Thesis/dissertation

Exploring Education Policies and their Socioeconomic Implications: Evidence and Insights from Juvenile Crime, College Admissions, and Gender Bias

by

Tatiana A. Reyes Hinrichsen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Card, Chair
Professor Jesse Rothstein
Professor Christopher Walters

Summer 2023

Exploring Education Policies and their Socioeconomic Implications: Evidence and Insights from Juvenile Crime, College Admissions, and Gender Bias

Abstract

Exploring Education Policies and their Socioeconomic Implications: Evidence and Insights from Juvenile Crime, College Admissions, and Gender Bias

by

Tatiana A. Reyes Hinrichsen

Doctor of Philosophy in Economics

University of California, Berkeley

Professor David Card, Chair

This dissertation examines the different aspects of education policies and their socioeconomic implications through three distinct research papers. The first chapter investigates the causal impact of grade retention in primary school on juvenile crime in Chile and it was written jointly with Juan Diaz, Nicolas Grau, and Jorge Rivera. Utilizing a fuzzy regression discontinuity design, the study finds that repeating an early grade in primary school reduces the probability of committing a crime as a juvenile by 14.5 percentage points. By employing a dynamic model, the research demonstrates that the observed result is primarily driven by two mechanisms associated with the timing of grade retention. First, grade retention in early grades decreases the likelihood of subsequent retention in later grades. Second, late grade retention in primary education exerts a stronger positive effect on crime reduction compared to direct effects in early grades. The findings suggest that, if grade retention remains an ongoing policy, its optimal implementation at the margin should involve retaining students in early grades to prevent retention in later ones.

The second chapter explores the equity and efficiency effects of a 2013 reform in the Chilean college admissions system. This reform aimed to increase equity by introducing a third component based on the student's GPA relative to the historical average at their high school. By simulating the admission mechanism with and without the relative GPA boost, the study categorizes applicants into three groups: those who gained access to more selective programs (pulled-up), those who lost access to more selective programs (pushed-down), and those whose admission was unaffected. Employing a difference-in-differences design, the research estimates the impacts of the reform on enrollment, persistence, and graduation. Pulled-up students were able to persist in their newly accessed programs, leading to more selective degree attain-

ment without significant effects on overall BA completion. Pushed-down students, predominantly from better-educated/higher-income families, experienced comparable reductions in the probability of graduating from selective programs, offset by gains in graduation from less selective programs. The study concludes that the reform improved equity with minimal or no loss in efficiency.

The third chapter estimates gender bias in a college admissions system based on standardized test scores and was jointly written with Matias Grau, Nicolas Grau, and Damian Vergara. This research applies standard discrimination literature tools in the context of the Chilean centralized admission system. We show that marginally enrolled female applicants exhibit higher first-year grades and are more likely to graduate on time compared to marginally enrolled male applicants. By employing an outcome test model, the research translates outcome differences into disparities in selection thresholds. We interpret this result as gender bias against females who, being "equally qualified" - based on their latent college performance - are not admitted into the program. Approximately 10% of enrolled male students would not have been admitted to their programs under the female effective selection thresholds. A counterfactual exercise simulating the college assignment algorithm with gender-targeted test-score inflation to correct for estimated bias suggests that such a policy would enable an improvement for 2% of female applicants in their preferred options relative to the observed assignment. Similar size effects are found with a decrease in the test score weight assigned to the application score formula.

Collectively, these research papers shed light on the complex relationships between education policies, socioeconomic outcomes, and student trajectories, emphasizing the importance of evidence-based policy recommendations to promote equitable educational systems.

Acknowledgments

I am deeply appreciative of the privilege of having David Card as my principal advisor. His mentorship, kind and rigorous, extended beyond the realm of economics and formed an indelible mark on my academic journey. I also thank David for giving me the opportunity to work with Laura Giuliano; both deeply shaped my PhD experience and research in the best way possible and I greatly admire their intelligence, ethical compass, and strong commitment. I extend my thanks to Jesse Rothstein and Christopher Walters for inspiring and teaching me about Labor Economics, Economics of Education, and Applied Econometrics. Both are brilliant and generous, and I am grateful to them. For their good advice in the moments I needed it, I would like to thank Pat Kline, Demian Pouzo, and Enrico Moretti.

I am also grateful to my coauthors and friends. I thank Pablo Munoz, Amanda Loyola, and Alonso Bucarey for setting an example for me to follow; I thank Damian Vergara for his infinite patience and willingness to teach and help everyone during the Ph.D.; and I thank Nicolas Grau, Juan Diaz, Jorge Rivera, and Matias Grau for their work and support. I consider myself lucky for having such bright coauthors.

Many colleagues and friends also deserve recognition. I thank Jimmy, Alfredo, Mathieu, Simón, Pia, Sebastian, Catalina, Elisa, Antonia, Jose, Harry, Livia, Daniela, Kayleigh, Cristobal, Felipe, Arlen, Bernardita, Camila, Sofia, Valentina, Francisca, Rocio, Belen, and Paloma for providing friendship, advice, inspiration, and support to me when I needed it. Financial assistance from the doctoral scholarship Becas Chile, and the Institute for Research on Labor and Employment allowed me to finance research travels to gather data and to work in secure computers located at government facilities. These scholarships improved the quality of my life significantly. I thank these entities and I thank my advisor for offering me research and teaching assistant jobs when I needed them.

Finally, I will be eternally grateful to my family. I thank my sisters and my parents for setting the example of hard work and loving life. I thank all the students I have thought; they motivated my research. I dedicate my dissertation to them.

Contents

List of Figures

List of Tables

Chapter 1

The impact of grade retention on juvenile crime

1.1 Introduction

Does grade retention in school increase the likelihood that young people will engage in criminal activity? From an opportunity cost point of view, it seems reasonable to argue that students may be more prone to pursue non-educational activities if they are not promoted to the next grade (Lochner). Conversely, repeating a grade might strengthen knowledge and improve discipline with potentially positive effects on a particular student's outcomes. Thus, instead of representing a "cost" for students, grade retention could be viewed as an "opportunity" that may help a young person become more competitive in the classroom and discourage divergence to non-educational activities (Jacob). The latter scenario is particularly relevant if early grade progression for students at the margin of minimum learning requirements can increase the probability of grade retention in the future. The ambiguity surrounding the potential effect of grade retention on crime is at the core of the well-known controversy relating to grade retention[1]

Settling this issue in an empirical way is particularly important in developing countries where the rates of both grade repetition and juvenile crime are much higher than those observed in developed countries. In 2012 the average rate of grade retention in primary education was 5.1% in developing countries but 1.4% in developed countries (UNESCO Institute for Statistics). In Chile, the average rate of grade

[1]The grade retention controversy exists because of ambiguous, and even contradictory, evidence regarding the effect that this measure has on some academic and socio-emotional outcomes of students, see Holmes et al. and Jimerson; see also Reschly and Christenson for a fresh look at this subject.

retention, although below the average for developing countries, has been increasing over the last 20 years, rising from 3.1% to 3.8% between 1999 and 2012. Regarding crime levels in a broad sense, Chile has a higher incarceration rate than OECD countries, detaining 266 inmates per $100,000$ as opposed to 145.5.[2]

Despite the vast literature linking grade retention and juvenile crime,[3] evidence of a causal effect between them is scarce and, as discussed later in this section, does not exist for developing economies. One factor that may explain the lack of evidence is the difficulty in finding an adequate empirical setting and dataset to overcome the potential endogeneity produced by the fact that the latent outcome —namely, criminal activity that would be observed in the absence of grade retention —and the propensity to fail a grade are simultaneously determined.

To fill the gap, this paper estimates the causal effect of grade retention on juvenile crime using a regression discontinuity (RD) approach. More specifically, we rely on the discontinuity in the probability of grade retention generated by the most commonly applied rule used to determine grade retention decisions in Chile —namely, that the grade must be repeated when a student scores below 4 in two or more subjects and has an average score lower than 4.95 across all subjects.[4] Although students can be retained due to more than one rule, the conditions of the rule selected for this paper were fulfilled in 84% of the cases of total grade retentions in Chile in 2007 (the year considered in our estimation sample).

We use an exceptional database from Chile, which matches individual academic records for all students (1st to 12th grade) with youth and adult criminal prosecution information, also on an individual basis, between 2007 and 2019. Our estimation sample includes students who attended 2nd and 3rd grade in 2007 and had not previously been the subject of grade retention. We restrict the sample to 2007 because it is the earliest year we can observe the annual average score for all subjects taken by students, which is required to evaluate the implementation of the retention rule. Moreover, we focus our attention on early grades because we see no evidence of manipulation in terms of grade retention decisions (i.e., sorting around the threshold for grade retention).[5]

---

[2]European Institute for Crime Prevention and Control, affiliated with the United Nations.

[3]For instance, Burdick-Will, Fagan and Pabon, and Hirschfield, among others, have shown how criminal activities affect some schooling outcomes, a sort of the inverse of the problem studied here. The effect of compulsory schooling laws on crime has been investigated by Lochner and Moretti, Brugård and Torberg, and Machin, Marie, and Vujić, among others. Other contributions have investigated how school starting age may affect crime (see Landersø, Nielsen, and Simonsen and references therein).

[4]For reference, the potential scores range from 1 to 7 (awarded in increments of 0.1).

[5]The 1st grade is not considered because Chilean law gives much more agency to teachers on

Our main finding is the robust evidence of a (local) negative causal effect of grade retention on juvenile crime. We implement the standard fuzzy RD procedure developed by Calonico, Cattaneo, and Titiunik and Calonico et al. We find that repeating a grade decreases the probability of committing a crime as a juvenile by 14.5 percentage points (pp) and by 10.7 pp in the case of a severe crime. Reassuringly, the results are not sensitive to the bandwidth choice or the implementation of higher-order polynomials; we also find no effect when we use placebo cutoff values. Additionally, we examine the effect of grade retention on dropping out of school, finding that grade retention in 2nd or 3rd grade decreases the probability of dropping out by 31 pp.[6]

Given that we also find that retention during the early grades improves future grade point average (GPA) and decreases the probability of future grade retention, we complement the RD analysis by estimating a (semi-structural) dynamic model.[7] In this model, grade retention in the early grades can directly and indirectly (via future GPA and future grade retention) affect crime. By estimating this model we show that the results from our RD estimation are not driven by a direct and relevant negative effect of grade retention on crime, but they are driven by a combination of a negative effect of grade retention in early primary grades on grade retention in late primary grades, with an increasing impact of grade retention on crime as students progress through the primary grades. Thus, our findings support the idea that conditional on the decision to keep grade retention as an ongoing policy, the best implementation of this policy for those students around the threshold of the retention rule is to be retained in early grades in order to avoid retention in later ones.[8]

There is a growing literature that examines test-based promotion policies on measures of academic performance and crime. Greene and Winters, looking at 3rd-grade students in Florida, show that retained students slightly outperformed students that were socially promoted (i.e., who should have been retained under the policy in place at the time). Jacob and Lefgren, looking at students from Chicago, find a differential effect between grade retention in the 6th and 8th grades. Their results show a substantial increase in the probability of dropping out of high school if a student is subject to retention in the 8th grade. To the best of our knowledge, there are two papers more closely related to our investigation —namely, Eren, Depew, and Barnes and Eren, Lovenheim, and Mocan. Both estimate the impact of grade retention

grade retention decisions at this level.

[6]The issues related to grade retention and school dropout have been investigated by Roderick and Manacorda. See King, Orazem, and Paterno for a comprehensive literature review.

[7]Future GPA is defined as a student's average GPA between 4th and 8th grade.

[8]There are many other educational policies that can be useful to prevent juvenile crime. Given what we can learn from our model, here we only focus on the timing of the grade retention.

(after the offer of a summer school program) on juvenile and adult delinquency (as well as other outcomes) in Louisiana. They assemble a novel dataset after merging administrative information on educational outcomes with the criminal records of students attending schools in Louisiana. Then, taking advantage of the test-based grade promotion policy, the authors build a fuzzy RD design where the forcing variable is a standardized test score that determines whether or not a student is promoted. Eren, Depew, and Barnes conclude that there is no effect of this test-based grade retention policy for 4th-grade students; for students attending the 8th grade, the policy has a small negative impact at most. Eren, Lovenheim, and Mocan show that when looking at a longer period (i.e., criminal convictions until age 25) being retained in the 8th grade has large effects on the likelihood of being convicted of a crime (and the number of convictions). The results presented by Eren, Lovenheim, and Mocan is consistent with our results as well as other evidence showing that the effect of grade retention varies as a function of when students are retained (Ou and Reynolds, Fruehwirth, Navarro, and Takahashi). Our dynamic model addresses the differing results from the literature related to the impact of grade retention on academic performance and crime by stressing the crucial role of grade retention timing as well as the impact of the treatment on the probability of being treated in the future.

A similar study was carried out by Cook and Kang who merge administrative data for academic performance with the criminal records of students attending public schools in North Carolina. They exploit a sharp RD design generated by the specific date that establishes the minimum age for school enrollment ("the cut date") and assess its effect on a number of educational outcomes as well as on crime committed as a juvenile. They highlight two main findings. First, middle school students born just after the cut date (i.e., the oldest children in each grade) are more likely to outperform (in mathematics and reading) those born just before the cut date (i.e., the youngest children in each grade), and the oldest children are also less likely to be involved in juvenile delinquency. Second, those children born just before the cut date are more likely to drop out of school and commit a severe offense. Finally, Depew and Eren, exploiting the same discontinuity as in Eren, Depew, and Barnes. and using the same student data from Louisiana, find that delaying school entry by one year decreases the frequency of juvenile delinquency for young black females.

In the context of the existing literature, this paper makes two main contributions. To the best of our knowledge, this is the first paper that studies the causal effect of grade retention on crime in a developing country. To the extent that grade retention and juvenile crime are much more prevalent in these countries, this is a relevant contribution. Moreover, by estimating a dynamic model, we show that our finding of a negative effect of grade retention on crime, which is also present in the literature, can be explained by a dynamic effect of grade retention in early primary grades on the probability of grade retention in late primary grades. In other words, it is not due to a direct negative effect of grade retention on crime. As we discuss in the conclusion, this result may be useful in order to better understand the heterogeneity in the results observed in the literature and, hence, it may be very relevant for policy design.

This paper is organized as follows: In Section 2 we describe the institutional background and the main features of both the educational and criminal datasets, and we present the evidence regarding the discontinuity created by the retention rule. In Section 3 we present our empirical strategy and study its validity. In Section 4 we outline our main results. Section 5 introduces the dynamic model and discusses its simulation results. And, finally, Section 6 concludes.

1.2 Institutional background, data, and retention rules

In this section, we describe the Chilean school system and juvenile criminal justice system, then we describe the characteristics of our dataset. Finally, we explain how the grade retention rule operates, which is critical in order to understand the source of the exogenous variation in our empirical strategy.

School system

In Chile, primary education lasts from the 1st to the 8th grade, and secondary education from the 9th to the 12th grade.[9] Primary education has a unified curriculum that consists of a set of minimum subjects; in order to progress to the next grade, a student must attain a certain level of academic knowledge. The Ministry of Education provides guidelines for grade retention. The guidelines state that a student should be retained if their GPA or attendance rate falls below a certain level (described below). In this paper, we study the effect of grade retention at 2nd or 3rd grade on juvenile crime. We select these particular grades for study because in later grades there is evidence of scores manipulation (see Solis), which raises questions about the utilization of a regression discontinuity approach, and because the law gives more agency to teachers regarding grade retention decisions at the first-grade level.[10]

[9]Secondary education (or until age 21) became mandatory in 2003.

[10]The law establishes that the school principal in concordance with the academic coordinator and the teaching staff (after their consultation) can waiver the attendance condition. Additionally, if the student does not fulfill the requirements to be promoted, or fails an important subject, the

Each grade in primary education is comprised of approximately 250,000 students who can attend public, private subsidized, or private unsubsidized schools. The first two types of school account for 93% of the total enrollment (in similar proportions) and receive the same per-student subsidy.[11] In this paper we consider students from all school types.

Dropout in Chile is low compared with Latin-American countries or the US. According to official statistics for 2012, it was 3.7% and it dropped to 2.2% in 2019. In Chile, most dropout happens before completing secondary education and towards the beginning of high school. In our dataset, we define dropout as one when the student does not enroll in 12th grade until 3 years after his/her expected graduation date. This definition overestimates dropout because a student who was enrolled in 2nd grade may not be enrolled until 12th grade for several reasons: migration, change to alternative types of education like adult education, and educational lag bigger than three years, among others. There are no reasons to expect this measurement error to be non-randomly distributed in our population. However, all things considered, the analysis and quantitative interpretation of the dropout results should be taken with caution.

To measure the learning process, knowledge acquisition, and school performance of students there is a system of national standardized testing (SIMCE) in which all students in 4th grade must participate.[12] The government uses the results from the SIMCE tests to allocate resources and inform the public about the quality of schools by listing school-level results in major newspapers. Since all schools, including public schools, are funded on the basis of a per-student formula there is significant pressure to produce good SIMCE test results. This, in turn, creates an impetus for selecting and expelling students as well as for increasing grade retention in order to improve academic performance.

Juvenile criminal justice system

The juvenile criminal justice system in Chile was reformed in 2005 (Act N^o 20084) and came into effect in 2007. Inspired by the United Nations Convention on the Rights of the Child, it is based on the principles of an exceptional and moderate application of criminal law and the use of confinement only as *ultima ratio* (see

school has to evaluate the reasons and social and emotional condition of the student. They would need to make the case, with all the relevant documentation, to proceed in a different way than what the rule requires.

[11] More details about Chilean education system can be found in Gauri and Grau, Hojman, and Mizala.

[12] See Meckes and Carrasco for details.

Langer and Lillo). This reform made three major changes to the previous system. It reduced the age of criminal responsibility from 16 to 14. It ended the ambiguity of the previous system whereby adolescents could be treated as adults or juveniles depending on the considerations of the judge. And, for convicted juvenile defendants, it reduced the punishment by one grade relative to the corresponding adult sentence.[13] Furthermore, the new juvenile criminal justice system was implemented as Chile was undertaking a radical reform of its criminal justice system as a whole, which began in 2000 and was completed in 2005. This broad reform replaced the inquisitorial model, a written system that had been in place for more than a century, with an oral, public, and adversarial procedure.[14] As part of the reform, several new institutions were created including the Public Defender's Office (PDO) and the Public Prosecutor's Office. The PDO provides free legal representation to almost all individuals who have been accused of committing a crime, and it collects information on all defendants that use their services, both juveniles and adults, which includes detailed information on the particular crime in question. Our data on juvenile criminal activity comes from PDO records.

In this paper, we measure crime, our dependent variable, as being prosecuted. We consider two types of crimes: *all crimes*, an indicator variable that takes the value of one when the juvenile was prosecuted during the ten years following 2007 (the year that defines treatment);[15] and *severe crimes*, an indicator variable that takes the value of one when the juvenile was prosecuted for a severe crime during the period already described. Following previous literature (see Cortés, Grau, and Rivera), we define severe crime as a type of crime for which the pretrial detention rate is greater than 3%. Table A.1 (Appendix A.1) presents the distribution of juvenile crimes across different types of crime. Figure A.1 (Appendix A.1) shows, among repeaters and non-repeaters in early primary school grades, the fraction of students prosecuted for the first time at different ages.

Data

We assemble our administrative dataset using data from the Ministry of Education and the PDO. For youths not legally represented by a PDO attorney (i.e., they have a private attorney), we observe the alleged crime but we do not observe the final

[13]See Couso and Duce for a detailed description of this reform.

[14]See Blanco, Hutt, and Rojas for a detailed description of the criminal justice system reform in Chile.

[15]Because our estimation sample is comprised of students with different ages (i.e., we consider 2nd and 3rd-grade students), we follow those students who attended 2nd grade in 2007 for one more year.

verdict. That said, less than 3% of prosecuted youths in our dataset are represented by a private attorney. In this paper, we use PDO records for juvenile criminal cases prosecuted during the period between January 2008 and December 2018.

The information collected from the Ministry of Education is an administrative panel dataset for every student in Chile between 2002 and 2019. The dataset indicates the school attended each year, the grade level (and whether the grade was repeated), the student's attendance rate, some basic demographic information, and (for 2007 only) the student's annual average score for each subject (cumulative GPA).[16] The cumulative GPA for each subject is critical information in the context of our RD approach because it is needed to build a more continuous measurement for the average across all subjects.[17] From this panel we build the other dependent variables considered in this paper: future grade retention, defined as at least one retention between 4th and 8th grade; future GPA, defined as a student's average GPA between 4th and 8th grade; and dropout. The latter is defined as permanently absent without graduation from 12th grade. We merge this panel with the data from the SIMCE test, which is taken annually by all 4th-grade students. When students take the SIMCE test a survey is administered to their parents. From these surveys, we obtain information about both parents' education level and family income.

Retention rules and discontinuity

In Chilean primary education students are taught around 10 subjects per year and are scored between 1 and 7 for each subject with an increment of 0.1. In this context, there are three important rules for grade retention. Students have the right to progress to the next grade, unless: (1) their attendance rate is below 85%; (2) they score below 4 for one subject and have an average score across all subjects lower than 4.45; or (3) they score below 4 on two or more subjects and have an average score across all subjects lower than 4.95. In 2007 489,168 students attended the 2nd and 3rd grades (the two grades included in our estimation sample) and 20,309 repeated the grade. Of those students, 2,634 (13%) were retained after scoring above 4 in all the subjects (probably due to a low attendance rate), 548 (2.7%) repeated the grade after scoring below 4 in only one subject, and 17,127 (84.3%) were retained after scoring below 4 in two or more subjects. Although students can be retained

[16] To distinguish between two distinct uses of the word "grade" —that is, between a level and an academic performance —grade performance will, hereafter, be referred to as "score".

[17] For other years, we only have the average across all subjects officially reported by the Ministry of Education. The problem with this measurement is that it is approximated and, consequently, using this level of aggregation for our estimation would mean comparing students with an average of 4.4 with students with an average of 4.5, for example.

due to more than one rule, given the distribution of cases the most relevant is the last rule described, where the threshold is 4.95 and the condition to be applied is scoring below 4 in at least two subjects. Therefore, in this paper, we exploit the discontinuity of treatment probability around a GPA of 4.95 as exogenous variation in the probability of grade retention.[18]

Given our research question and the characteristics of the selected retention rule, the estimation sample has the following characteristics. We focus our attention on the students who attended the 2nd and 3rd grades in 2007. As stated, we restrict our sample to those grades because for later grades we observe some evidence of manipulation in grading decisions around the 4.95 threshold (for more detail regarding this manipulation see Solis) and because the law gives more agency to teachers on grade retention decisions at the first-grade level. For obvious reasons, we only consider students affected by the aforementioned retention rule —namely, those scoring below 4 in two or more subjects. In order to exclude schools where no student scores below the threshold, we only consider schools where at least one student scores less than 4.95 on average across all subjects. Finally, we focus on students who had not previously repeated a grade. In the last section of the paper we develop and estimate a dynamic model that, besides helping us understand our results, allows us to understand the effect of more than one grade retention.[19]

Estimation sample

The final dataset includes $13,072$ students. The overall impact of sample restrictions is observed in Table 1.1, which shows how the estimation sample is different from the full sample (i.e., the population) in terms of a set of observables. As can be observed, in terms of individual characteristics the estimation sample has a greater percentage of males and students from less educated families, and the students perform less well at school, including a greater grade retention rate. The schools in the estimation sample report lower average standardized test scores and lower average education for both parents. Furthermore, the students in the estimation sample have

[18]The school principal, in concordance with the academic coordinator and the teacher staff (after their consultation), can waiver the attendance condition. Additionally, if a student does not fulfill the requirements to be promoted, or fails an important subject, the school has to evaluate the reasons for this academic performance and the social and emotional condition of the student. They would need to make the case, with all the relevant documentation, to proceed in a different way with respect to the rules. This is a costly process. That said, we also present the reduced form estimations (i.e., the sharp RD estimation), which are not affected by this discretion.

[19]As our treatment is grade retention in 2007 we should exclude students who were prosecuted before 2008; however, the ages of students in our 2007 estimation sample means that this is not a binding restriction.

a higher probability of committing a crime in the future, approximately twice the probability of the population. The differences between the full and estimation sample demonstrate that by restricting our attention to students close to the retention cutoff, the estimation sample is comprised of low-performance students. This is also observed in the table. The differences also emphasize that, as is usually the case when the causal effect is estimated using an RD empirical strategy, our results and their causal interpretation only have local validity.

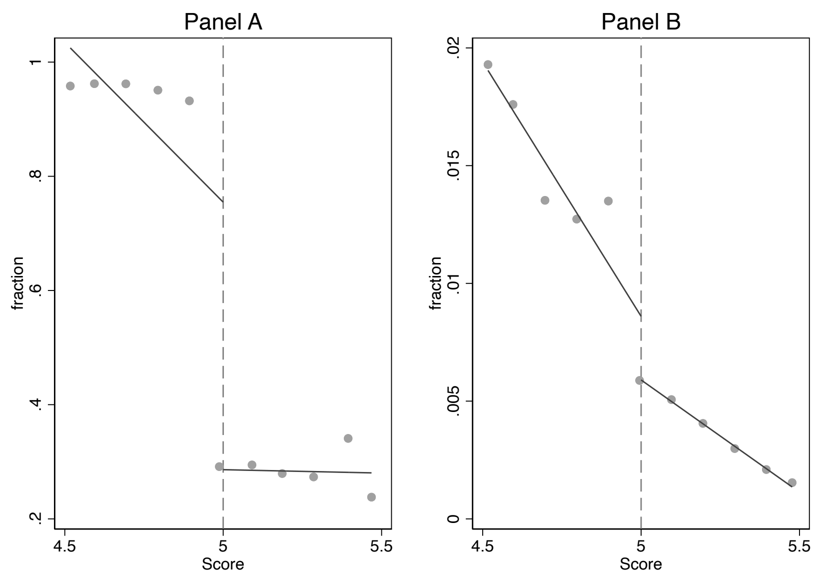
Table 1.1: Estimation sample versus full sample

	Full Sample [$n = 489,168$]		Estimation Sample [$n = 13,072$]	
Variables	Mean	Std. Dev.	Mean	Std. Dev.
Panel A. Individual Characteristics				
Male (%)	51.5%	50.0	60.1%	49.0
Father's Education	11.2	3.9	9.2	3.7
Mother's Education	11.0	3.7	9.1	3.6
Attendance (2006) (%)	93.1%	10.0	92.0%	6.6
GPA (2006)	6.1	0.7	5.2	0.5
Panel B. School Characteristics				
Grade Retention Rate (last 3 years)	0.05	0.07	0.06	0.06
Average Math Standardized Score	0.0	0.6	-0.2	0.5
Average Verbal Standardized Score	0.0	0.5	-0.2	0.4
Average Income Decile	5.4	2.0	4.6	1.5
Average Father's Education	11.0	2.6	9.9	2.0
Average Mother's Education	11.2	2.3	10.2	1.9
Average Expectation of Child Education	15.4	1.7	14.7	1.5
Panel C. Outcome & Other Variables				
All Crimes (%)	6.6%	24.9	13.0%	33.7
Severe Crime (%)	5.2%	22.3	10.4%	30.6
Future Grade Retention (%)	33.8%	47.3	64.5%	47.9
Dropout (%)	24.4%	43.0	50.1%	50.0
GPA From 4th to 8th Grade	5.6	0.5	5.2	0.4
GPA (2007)	59.1	6.7	44.2	4.1
Grade Retention (2007)	0.04	0.20	0.93	0.26

Notes: The estimation sample considers students who attended 2nd or 3rd grade in 2007, who scored below 4 in two or more subjects, from schools with at least one student who scored less than 4.95 on average across all subjects, and who had not been subject to previous grade retention. The full sample includes all the students who attended 2nd or 3rd grade in 2007.

To explore the discontinuity due to the aforementioned retention rule, panel (a) of Figure 1.1 presents the probability of grade retention around the 4.95 threshold for those students who belong to the estimation sample. The figure shows a discrete and relevant change in the probability around the threshold. More specifically, this probability increases by 61.7 pp for those individuals marginally below the 4.95 threshold. To show that the rule is only binding under its specific conditions, in panel (b) we show the same exercise only considering those students who score below 4 in one or no subjects. In these cases, the grade retention probability is continuous around the threshold.

Figure 1.1: First Stage



Notes: This figure presents a binscatter plot of the fraction of students in 2nd and 3rd grade not promoted in 2007, with a linear fit at each side of the cutoff. Panel A includes students from the estimation sample with two or more subjects below 4. Panel B includes students without two or more subjects below 4.

1.3 Empirical approach

Exploiting the retention rule outlined in the previous section, we estimate the effect of grade retention on juvenile crime using the method to implement a fuzzy RD approach developed by Calonico, Cattaneo, and Titiunik and Calonico et al. In this context, the bandwidth selection is calculated by minimizing an approximation to the asymptotic mean squared error of the point estimator and removing the bias due to the curvature of the regression function.[20]

[20]We implement this approach by using the Stata routines developed by Calonico, Cattaneo, and Titiunik (using updated code as of 2020).

Let n denote the number of students in the sample. For individual i, let Z_i be the GPA score in 2007, the running variable in our application, whose cutoff level is denoted by $\bar{z}$ (which in our scenario is 4.95); and let W_i be the treatment indicator that takes the value one if the ith student repeats the grade in 2007; and let Y_i be the binary outcome that takes the value of one when the individual committed a crime as a juvenile and zero otherwise. Finally, X_i is a set of covariates. Given an optimal bandwidth h we calculate the RD estimation, τ_{FRD}, as:

$$\widehat{\tau}_{FRD}(h) = \frac{\widehat{\tau}_Y(h)}{\widehat{\tau}_W(h)},$$

$$\widehat{\tau}_Y(h) = \widehat{\alpha}_{Y,-}(h) - \widehat{\alpha}_{Y,+}(h), \qquad \widehat{\tau}_W(h) = \widehat{\alpha}_{W,-}(h) - \widehat{\alpha}_{W,+}(h),$$

where, for $J = Y, W$, the estimators $\widehat{\alpha}_{J,-}$ and $\widehat{\alpha}_{J,+}$ come from a standard local linear RD estimator:

$$\begin{pmatrix} \hat{\alpha}_{J,-} \\ \hat{\alpha}_{J,+} \\ \hat{\beta}_{J,-} \\ \hat{\beta}_{J,+} \\ \hat{\gamma}_J \end{pmatrix} =_{\alpha_{J,-},\alpha_{J,+},\beta_{J,-},\beta_{J,+},\gamma_J} \sum_{i=1}^{n} [J_i - 1(Z_i < \bar{z}) \cdot (\alpha_{J,-} + \beta_{J,-} \cdot (Z_i - \bar{z})) -$$

$$1(Z_i \geq \bar{z}) \cdot (\alpha_{J,+} + \beta_{J,+} \cdot (Z_i - \bar{z})) - \gamma_J \cdot X_i]^2 \cdot \frac{K\left(\frac{Z_i - \bar{z}}{h}\right)}{h},$$

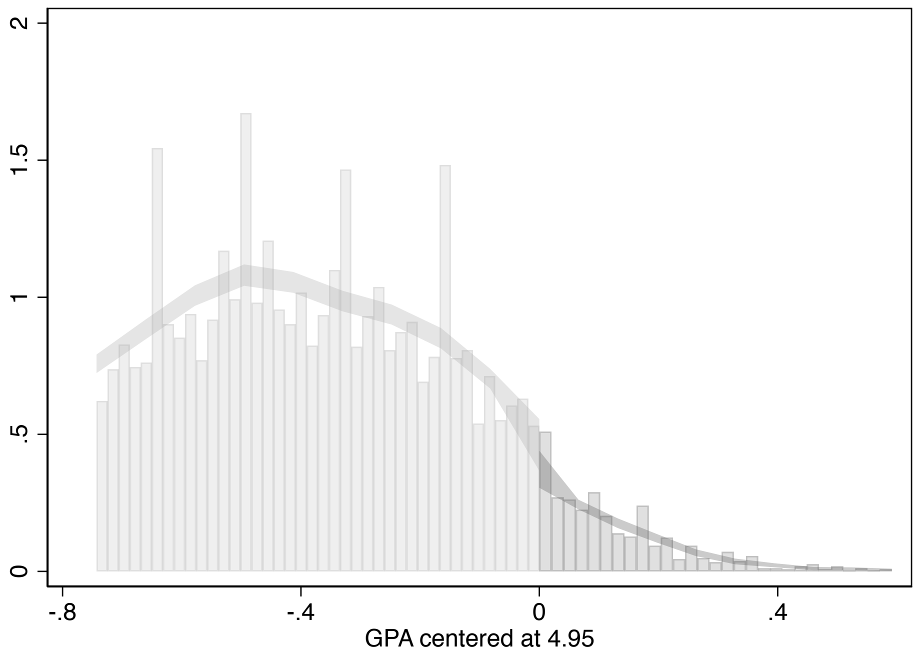
where $K(\cdot)$ is a kernel function. We cluster errors at the school level.

Validity of the RD Design

We explore the validity of the RD design by performing the two most common tests for this purpose: (i) we study the continuity of the density of the running variable at the cutoff and (ii) we examine whether covariates (i.e., observed variables measured before 2007) are similar between estimation sample students who are below and above the cutoff.

For (i) we implement the density test developed by Cattaneo, Jansson, and Ma. In concrete terms, we assess whether the density of the GPA (the running variable in our application) is a continuous function at 4.95 (the cutoff). As shown in Figure 1.2, the test result reveals that the null hypothesis stating that the density of the GPA is a smooth function at 4.95 cannot be rejected with a value of the robust bias-corrected statistic of 1.096. This implies a p-value of 0.27. This finding provides evidence in favor of the validity of our RD design because it suggests that the GPA is not determined by strategic behavior or manipulation at the cutoff.

Figure 1.2: Density test



Notes: The plot shows the density test proposed by Cattaneo, Jansson, and Ma. We implement this test using the Stata command rddensity. The value of the robust bias-corrected statistic of this test is equal to 1.096. This implies that we cannot reject the null hypothesis of continuity of the density function of the running variable at the cutoff (p-value equal to 0.27).

For (ii) we estimate the effect of being marginally below the cutoff on several covariates for students at the cutoff. Specifically, we consider 12 covariates in total, which are related to students' academic achievements and socio-economic backgrounds as well as certain characteristics of the students' schools.

As can be observed in Table 1.2, all covariates are similar between students who are marginally below and above the cutoff, except in the case of 2006 GPA where the difference between the two groups is statistically significant. Importantly, the significant difference between the two groups at the cutoff found in 1 of the 12 covariates can be explained by chance in a setting of multiple comparisons rather than by a systematic difference between the two groups. Given that students who are marginally below and above the cutoff are similar in covariates, these results provide evidence that supports the validity of our RD design as any difference in the observed outcome of interest (juvenile crime, dropout, or future grade retention) between the two groups of students at the cutoff can be attributed to the treatment (i.e., being retained in 2007).

Table 1.2: Differences in covariates at the cutoff for retention

Variable	RD Estimator	Robust Inference p-value	Robust Inference C.I.	Number of Observations
Panel A. Individual Characteristics				
Male	0.07	.4	[-0.09 0.24]	13,029
Father's Education	0.39	.59	[-1.37 2.40]	7,981
Mother's Education	0.45	.46	[-1.08 2.37]	8,331
Attendance (2006)	1.26	.12	[-0.45 3.68]	12,567
GPA (2006)	0.18	.01	[0.04 0.36]	12,567
Panel B. School Characteristics				
Grade Retention Rate (last 3 years)	0.00	.41	[-0.03 0.01]	12,889
Average Math Standardized Score	-0.03	.94	[-0.21 0.19]	12,815
Average Verbal Standardized Score	0.06	.48	[-0.12 0.24]	12,817
Average Income Decile	0.23	.22	[-0.19 0.80]	11,262
Average Father's Education	0.06	.64	[-0.53 0.87]	11,280
Average Mother's Education	0.30	.15	[-0.18 1.14]	11,281
Average Expectation of Child Education	0.14	.34	[-0.26 0.76]	11,271

Notes: The table presents the results based on the methods for estimation and inference of a sharp RD developed by Calonico, Cattaneo, and Titiunik and Calonico et al. The model is estimated without covariates, and the dependent variables are the 12 covariates presented in the first column. Differences in sample sizes are due to missing values for some of the covariates. In an ideal RD context, all the point estimates should not be statistically significant. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator.

1.4 Results

In this section, we present our findings on the impact of grade retention on juvenile crime, juvenile severe crime, and dropping out of school. These results are based on the methods for estimation, inference, and bandwidth selection for fuzzy RD designs developed by Calonico, Cattaneo, and Titiunik and Calonico et al. We also provide several robustness analyses that reinforce the validity of our RD design and, therefore, the plausibility of our findings.

Impact on crime, severe crime, and dropout

In Table 1.3, we present our estimations for the impact of grade retention on juvenile crime considering all crimes and severe crimes only, and dropout, including and not including covariates in the estimation. The covariates considered include attendance rate in 2006, GPA in 2006, gender, school characteristics, grade dummies, school grade retention rate in the previous three years, school average SIMCE score for mathematics, school average score for the SIMCE verbal test, school average years of education for both the father and mother, school average family income decile, and school average expectation of childhood education. All the estimations are presented with errors clustered at a school level. As seen in Table 1.3, we estimate that repeating a grade in 2007 decreases the probability of committing a crime as a juvenile by 14.5 pp when we do not include covariates and 17.7 pp when we include them. In the case of severe crimes, the effect of repeating a grade is -10.7 pp when we do not include covariates and -13.3 pp when we include them. All these effects are statistically significant at the 1% and 5% levels for all crimes and severe crimes, respectively.

Table 1.3: Effect of grade retention on juvenile crime and dropout

	All Crimes		Severe Crime		Dropout	
	Without Covs. (1)	Including Covs. (2)	Without Covs. (3)	Including Covs. (4)	Without Covs. (5)	Including Covs. (6)
RD Estimator	-.144***	-.176***	-.107**	-.133**	-.306***	-.408***
	(.045)	(.045)	(.041)	(.043)	(.076)	(.078)
Mean Variable	.132		.106		.508	
Std. Dev. Variable	.339		.308		.5	
Robust Inference						
p-value	0.01	0.00	0.03	0.02	0.00	0.00
C.I.	[-.249 -.038]	[-.293 -.059]	[-.205 -.009]	[-.246 -.02]	[-.485 -.127]	[-.621 -.196]
Effective Obs.						
Left	1,611	1,356	1,425	1,200	1,612	1,357
Right	567	483	537	457	567	483
Optimal Bandwidth[a]	.168	.168	.161	.161	.171	.171

Notes: This table presents the results for the impact of grade retention on juvenile crime and dropout, based on the methods for estimation and inference for fuzzy RD designs developed by Calonico, Cattaneo, and Titiunik and Calonico et al. The reported means and standard deviations are control group statistics. The specifications (2), (4), and (6) include the following covariates: attendance rate in 2006, GPA in 2006, gender, school characteristics, and grade dummies. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses and clustered at the school level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

There are two aspects that must be stressed regarding these results. First, these are sizable effects. For the control group, the crime rate is 13% for all crimes and

10.4% for severe crimes only. This means that the effect of grade retention on crime (without covariates) represents a decrease of 112%, whereas for severe crime the effect corresponds to a decrease of 102%. Second, although the effects are a little higher in absolute value and estimated with basically the same precision when we include covariates, the results are reasonably stable with the inclusion of additional control variables, which is consistent with the evidence shown in the previous section on covariates balance at the cutoff.

Given the richness of our panel dataset, we can also examine the effect of grade retention on other outcomes. This allows us to present a more complete picture of what happens to the students' trajectories after they repeat the 2nd or 3rd grade. Specifically, for the probability of dropping out of school, we find that not being promoted to the next grade in 2007 decreases the probability of dropping out of school by 31.1 pp without covariates and by 41.1 pp with covariates. All these effects are statistically significant at the 1% level. To assess the size of these effects, it should be noted that 50.1% of the control group drop out of school at some point. It is remarkable, again, how stable the point estimates are to the inclusion of covariates in the estimation. And it is this stability that reinforces our confidence in the validity of the RD approach in this context.[21]

To present the RD results graphically, in Figure 1.3 we show the outcome values and an estimation of the regression functions via local linear regressions around the threshold of all crimes and dropping out of school. As can be seen in Figure 1.3 in each plot there is a jump in the outcome variable at the cutoff for grade retention (at zero), which reinforces the plausibility of the findings presented in Table 1.3.[22]

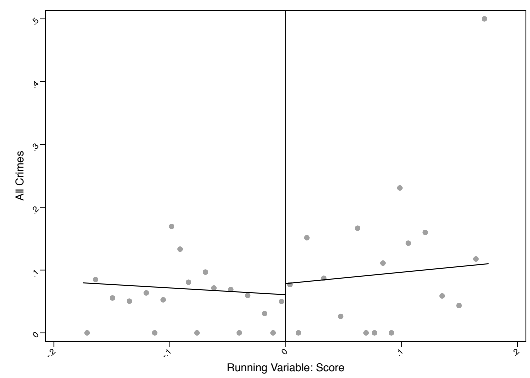
Robustness Analysis

We begin our robustness check by studying the extent to which point estimates are sensitive to the bandwidth choice. In our main estimation, following Calonico, Cattaneo, and Titiunik, this is calculated by minimizing an approximation to the asymptotic mean squared error of the point estimator and removing the bias due to the curvature of the regression functions. In Figure 1.4, we show our estimates for all crimes and severe crimes considering five possible bandwidths used in our main specification, with the middle bandwidth being the optimal value. These figures show that the estimation results regarding crime are not sensitive to bandwidth choice. Moreover, Figure A.3 (Appendix A.1) shows that the dropout point estimate is also not sensitive to bandwidth choice.

[21]Table A.2 (Appendix A.1) presents the reduced form estimations (i.e., sharp RD), for all the dependent variables analyzed in this paper.

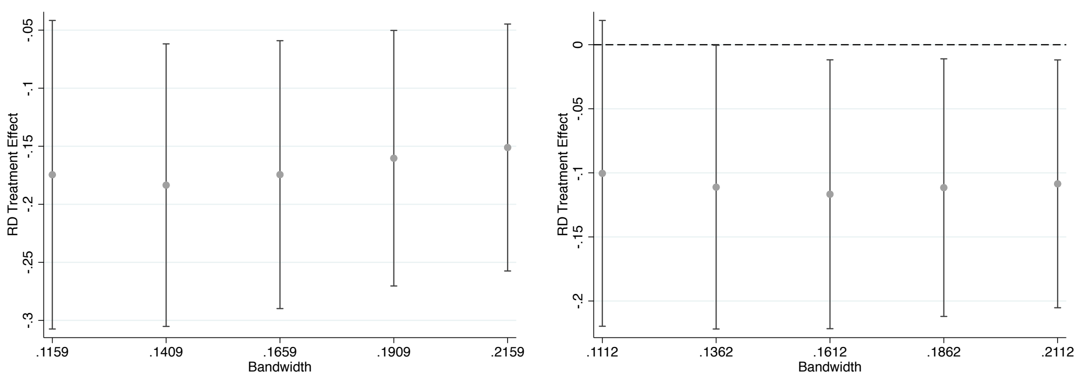
[22]Figure A.2 (Appendix A.1) shows the same plot but for severe crimes.

Figure 1.3: Graphic results for crime and dropout



Notes: This figure shows the outcome values and an estimation of the regression functions via local linear regressions around the threshold for grade retention for all crimes and dropping out of school.

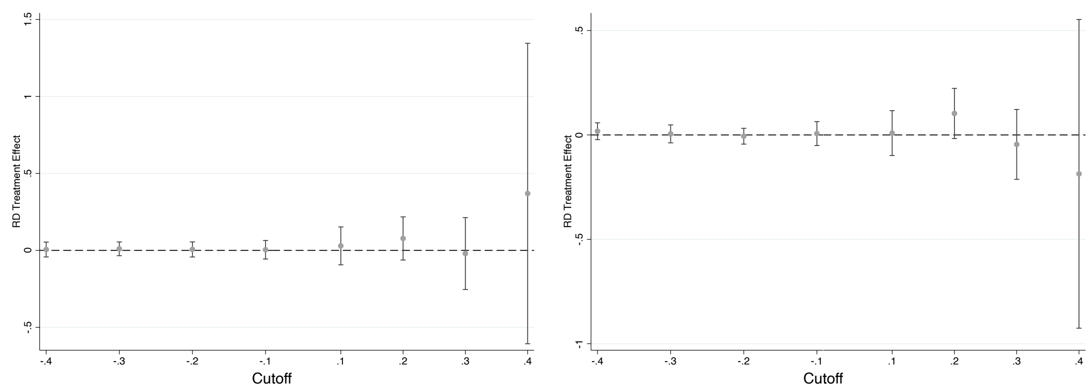
Figure 1.4: Sensitivity to bandwidth: all crimes and severe crimes



Notes: This figure shows the fuzzy RD estimations for the impact of grade retention on juvenile crime (using the methods developed by Calonico et al.), considering five different values of the bandwidth (the middle estimate is the optimal bandwidth). The point estimates are the dots and the confidence intervals at 95% are the brackets.

Our second robustness analysis performs estimation and inference of treatment effects in our RD setting but uses artificial (or placebo) cutoff values. Naturally, if our design is valid, we would expect that no significant treatment effects should appear at any artificial cutoff. In Figure 1.5 we present the results of the estimation, for all crimes and severe crimes, of the impact of being below the artificial cutoff considering eight possible artificial cutoffs, four that are below and four that are above the cutoff involved in the grade retention rule (i.e., 4.95).[23] As expected, these figures show that the effects are not statistically significant regardless of the artificial cutoff employed. For the case of dropping out of school, the effects are also not significant for distinct choices of placebo cutoffs, as can be observed in Figure A.4 (Appendix A.1). These results reinforce the validity of our RD design.

Figure 1.5: Placebo tests: RD estimations considering other cutoffs



Note: This figure shows the sharp RD estimations for the impact of being below the cutoff for all crimes and severe crimes (following the method developed by Calonico et al.), for cutoff values that do not have consequences in terms of grade retention probability. Since the original cutoff is at 4.95, we present four cutoffs below and four above 4.95, such that, −0.1 denotes 4.85 and 0.1 denotes 5.05. The point estimates are the dots and the confidence intervals at 95% are the brackets.

For our final robustness check we estimate the same fuzzy RD model as in the case of the main specification, but considering higher-order polynomials (up to the fifth order). In Table 1.4 we show the results from this exercise. In short, the point estimates considering higher-order polynomials (rows two to five) are similar to our main results (row one) across the different specifications and they are also statistically significant. Therefore, this exercise provides supporting evidence for the robustness of our results.

[23]Remember that the actual cutoff involved in the grade retention rule is 0 as we calculate it using the running variable centered at 4.95. We only consider eight artificial cutoffs, because the cutoff 4.45 is also relevant for grade retention probability.

Table 1.4: Effect of grade retention on juvenile crime and dropout: higher order polynomials

	All Crimes		Severe Crimes		Drop out	
Order of polyn.	est.	(s.e.)	est.	(s.e.)	est.	(s.e.)
1	-.144	.045	-.107	.05	-.306	.091
2	-.192	.069	-.119	.055	-.433	.127
3	-.188	.068	-.132	.066	-.467	.138
4	-.205	.082	-.11	.074	-.498	.161
5	-.201	.085	-.116	.074	-.533	.166

Notes: This table presents the results for the impact of grade retention on juvenile crime and dropout, considering higher-order polynomials (up to the fifth order). It uses the estimation sample described in Table 1.1 and follows the methods for estimation, optimal bandwidth definition, and inference for fuzzy RD designs developed by Calonico, Cattaneo, and Titiunik and Calonico et al.

1.5 Dynamics

Even though our RD empirical strategy delivers local and causal estimates for the effect of grade retention on juvenile crime, the results are not conclusive regarding the effect on juvenile crime of eliminating grade retention, not even for the compliers. In other words, the problem is not just the local nature of our estimates. There are two important reasons for this. On the one hand, grade retention can be an incentive to increase student academic effort, something that is not captured by our empirical approach. On the other hand, grade retention today can reduce the probability of grade retention in the future. Thus, students who belong to the control group in our RD design can be part of the treatment group in the future (in later grades). And to the extent that there is heterogeneous effect of grade retention on juvenile crime across grades, the negative effect that we find using a RD estimation is also consistent with positive effects of grade retention on juvenile crime in both the present (early primary grade) and the future (late primary grade), but where the effect in the future is larger than the effect in the present. While we do not have the data to study concerns relating to incentives, this section is devoted to addressing concerns about the dynamics.

We start this analysis by documenting the effect of grade retention in the 2nd or 3rd grade (2007) on future GPA, defined as a student's average GPA between 4th and 8th grade, and on future grade retention, defined as an indicator variable that takes the value of one if a student is subject to at least one grade retention between 2008 and 2014. In practice, this analysis involves running the same model specification that produced the results presented in Table 1.5 but changing the dependent variable from crime to future GPA or grade retention. Table 1.5 also shows that grade retention increases future GPA by 0.3 points (0.86 standard deviations) and decreases the probability of future grade retention by 43.2 pp. As the table shows, these point estimates are quantitatively relevant, statistically significant and robust to the inclusion of covariates.

Table 1.5: Effect of grade retention on educational outcomes

	Future Grade Retention		GPA From 4th to 8th Grade	
	Without Covs. (1)	Including Covs. (2)	Without Covs. (3)	Including Covs. (4)
RD Estimator	-0.427***	-0.492***	0.030***	0.032***
	(.072)	(.074)	(.005)	(.005)
Mean Variable	.645		5.163	
Std. Dev. Variable	.479		.351	
Robust Inference				
p-value	0.00	0.00	0.00	0.00
C.I.	[-.602 -.253]	[-.708 -.277]	[.018 .043]	[.018 .047]
Effective Obs.				
Left	1,617	1,361	1,275	1,077
Right	569	485	485	416
Optimal Bandwidth[a]	.177	.177	.163	.163

Notes: This table presents the results for the impact of grade retention on education outcomes, based on the methods for estimation and inference for fuzzy RD designs developed Calonico et al. The specifications (2), (4), and (6) include the following covariates: attendance rate in 2006, GPA in 2006, gender, school characteristics, and grade dummies. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses and clustered at the school level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

This result confirms that —relative to treatment RD group —our control group has a greater probability of being treated in the future. Thus, once again using the discontinuity in the probability of grade retention in 2007, we develop a dynamic model that allows us to identify and estimate the effect of grade retention at different grades (i.e., in different years). The parameters of this model are identified under assumptions that are more demanding than those needed in the RD approach but are still reasonable. We estimate this model using the same estimation sample from our RD estimation —namely, students who were attending 2nd or 3rd grade in 2007.

Model setting

The model has three periods. The first period ($t = 1$) is the year 2007 and corresponds to the students attending 2nd or 3rd grade. The second period ($t = 2$) is the group of years between 2008 and 2015 and corresponds to the students attending a grade between the 2nd and the 8th. The particular grade depends on the year, the grade the students attended in 2007, and the number of grade retentions sustained. Finally, the third period ($t = 3$) corresponds to the students being aged between 14 and 17 years, when an individual can commit a crime and be punished as a juvenile (C).

An individual i is characterized by the vector X_i, $G3_i$ and τ_i, where X_i and $G3_i$ are observable by the econometrician and τ_i is an unobservable variable with finite support, $\tau_i \in \{1, 2, ..., K\}$ (i.e., unobserved types). In the estimation, we consider three unobserved types ($K = 3$). X includes gender, the average education of both parents and an indicator variable that takes the value of one when educational information regarding the parents is missing from the SIMCE survey. $G3$ is an indicator variable that takes the value of one if the student is attending the 3rd grade in 2007; it is included to capture the fact that the results can be different depending on the starting point. In the first two periods, student i attends schools j (which can be different between $t = 1$ and $t = 2$). The schools are characterized by the vector of characteristics W_j^t, which considers the average education of fathers and mothers of students at the school, the average scores for mathematics and Spanish from the SIMCE test, and an indicator variable for public schools. The academic performance at period t is characterized by GPA^t and the grade retention indicator R^t ($t \in \{1, 2\}$). GPA^2 is defined as GPA when a student repeated a grade for the first time during the second period or the student's lowest GPA during $t = 2$ if they were not subject to grade retention during this period. Along similar lines, and given that it is possible for a student to attend different schools during the second period, we define the second-period school as the one where the student repeated a grade for the first time during $t = 2$ or, if the student was not subject to grade retention during that period, the school where they had the lowest average score across all subjects.

The dynamic model is given by the following equations:

$$GPA_{ij}^1 = \alpha_{\tau_i}^1 + X_i\alpha_x^1 + G3_i\alpha_{gr3}^1 + W_j^1\alpha_w^1 + \varepsilon_{ij}^1. \tag{1.1}$$

$$GPA_{ij}^2 = \alpha_{\tau_i}^2 + X_i\alpha_x^2 + G3_i\alpha_{gr3}^1 + W_j^1\alpha_w^2 + R_i^1\alpha_{R,\tau_i}^2 + \varepsilon_{ij}^2. \tag{1.2}$$

$$R_i^1 = 1\Big(\gamma_{\tau_i}^1 + 1(GPA_{ij}^1 < 4.95)\gamma_1^1 + (GPA_{ij}^1 - 4.95)\gamma_2^1 + (GPA_{ij}^1 - 4.95)^2\gamma_3^1 + G3\gamma_4^1 \geq \eta_i^1\Big). \quad (1.3)$$

$$R_i^2 = 1\Big(\gamma_{\tau_i}^2 + R_i^1\gamma_1^2 + GPA_{ij}^2\gamma_3^2 + (GPA_{ij}^2)^2\gamma_4^2 + 1(GPA_{ij}^2 < 4.5)\gamma_5^2 + 1(GPA_{ij}^2 < 5)\gamma_6^2 + G3\gamma_7^2 \geq \eta_i^2\Big). \quad (1.4)$$

$$C_i = 1\Big(\beta_{\tau_i} + R_i^1\beta_{R1} + R_i^2\beta_{R2} + R_i^1R_i^2\beta_{RR} + G3_i\beta_{g3} + GPA_{ij}^1\beta_{G1} + GPA_{ij}^2\beta_{G2} + X_i\beta_X + W_j^1\beta_{W1} + W_j^2\beta_{W2} \geq \eta_i^3\Big). \quad (1.5)$$

We assume that all shocks are normally and independently distributed with mean zero and variances equal to $\sigma_{\varepsilon 1}^2$, $\sigma_{\varepsilon 2}^2$, $\sigma_{\eta 1}^2$, $\sigma_{\eta 2}^2$, and $\sigma_{\eta 3}^2$, respectively.

There are some features of this model and the identification of its parameters that are worth highlighting. First, although shocks are independent, the model allows for correlation across dependent variables, conditional on observables, through the unobserved heterogeneity (τ_i). This approach is similar to allowing correlation among shocks and has the advantage that the unobserved type can also accommodate heterogeneity in the impact of the covariates.[24] In fact, we allow for heterogeneity in the impact of grade retention on GPA in the second period (via α_{R,τ_i}^2). This heterogeneity is useful because the effect can be due to a positive or negative effect of R_i^1 on future academic performance or to GPA manipulation impacting grade retention probability. Yet, these two mechanisms cannot be separately identified. Second, given the evidence supporting our RD strategy, we assume there is no manipulation in the first period and as a consequence, γ_1^1 identifies the exogenous variation in R_i^1, which enables the identification of α_{R,τ_i}^2, γ_1^2, and β_{R1}. To the extent that unobserved type is equivalent to allowing for correlation across shocks and that the discontinuity

[24]For example, given two random variables $\tilde{\varepsilon}_i = \varepsilon_i + \beta_{\tau_i}$ and $\tilde{\eta}_i = \eta_i + \alpha_{\tau_i}$, where ε_i and η_i are independent and τ_i is an unobserved heterogeneity with two types: $\tau_i \in \{0, 1\}$ and $Pr(\tau_i = 1) = \pi$. Then, $Cov(\tilde{\varepsilon}_i, \tilde{\eta}_i) = \pi(1-\pi)(\beta_1 - \beta_0)(\alpha_1 - \alpha_0)$.

works as an instrumental variable, the causal effect of grade retention in the first period on juvenile crime and on grade retention in the second period are identified given the same exclusion restriction that supports identification in the case of a bivariate probit model (see Li, Poskitt, and Zhao and Han and Vytlacil). Third, there are two sources of exogenous variation for R_i^2: the discontinuity in the probability of R_i^1 (which impacts R_i^2) due to the grade retention rule in the first period (which is not manipulated); and the discontinuity in the probability of R_i^2 due to the grade retention rule in the second period. For the second period, we do not observe a perfectly continuous measure for the average score across all subjects (i.e., it has an increment of 0.1), but we include two indicator variables that take the value of one when the average score is below each threshold. In this period, the manipulation is captured by α^2_{R,τ_i}. Fourth, the sources of exogenous variation for R_i^1 and R_i^2 allow us to identify the parameters of interest: β_{R1}, β_{R2}, and β_{RR}.

We estimate this model considering two samples. The first (the *full estimation sample*) is the same sample used in the RD estimation (see Table 1.1). The second (the *restricted estimation sample*) is the sample that restricts the full estimation sample to those students who are effectively considered in the RD estimation —namely, those with an average score between $4.95 - 0.16$ and $4.95 + 0.16$ —where 0.16 is the RD optimal bandwidth. Although in practice both estimation samples deliver similar results, we prefer the second sample because it makes the exogenous nature of the variation in R_i^1 more reliable.

Estimation

Let Ω be the set of parameters to estimate, such that $\Omega = \{\alpha^1, \alpha^2, \gamma^1, \gamma^2, \beta, \sigma\}$, the likelihood contribution of individual i, whose unobserved type is τ_i, is equal to:[25]

[25] $Z^R_{i,1} = \gamma^1_{\tau_i} + 1(GPA^1_{ij} < 4.95)\gamma^1_1 + (GPA^1_{ij} - 4.95)\gamma^1_2 + (GPA^1_{ij} - 4.95)^2\gamma^1_3 + G3\gamma^1_4$; $Z^R_{i,2} = \gamma^2_{\tau_i} + R^1_i\gamma^2_1 + GPA^2_{ij}\gamma^2_3 + (GPA^2_{ij})^2\gamma^2_4 + 1(GPA^2_{ij} < 4.5)\gamma^2_5 + 1(GPA^2_{ij} < 5)\gamma^2_6 + G3\gamma^2_7$; and $Z^C_i = \beta_{\tau_i} + R^1_i\beta_{R1} + R^2_i\beta_{R2} + R^1_iR^2_i\beta_{RR} + G3_i\beta_{g3} + GPA^1_{ij}\beta_{G1} + GPA^2_{ij}\beta_{G2} + X_i\beta_X + W^1_j\beta_{W1} + W^2_j\beta_{W2}$.

$$L_i(\Omega|GPA,X,G3,W,R,C;\tau_i)=\phi\left(\frac{GPA^1_{ij}-\alpha^1_{\tau_i}-X_i\alpha^1_x-G3_i\alpha^1_{gr3}-W^1_j\alpha^1_w}{\sigma_{\varepsilon 1}}\right)\frac{1}{\sigma_{\varepsilon 1}}$$
$$\phi\left(\frac{GPA^2_{ij}-\alpha^2_{\tau_i}-X_i\alpha^2_x-G3_i\alpha^2_{gr3}-W^1_j\alpha^2_w-R^1_i\alpha^2_{R,\tau_i}}{\sigma_{\varepsilon 2}}\right)\frac{1}{\sigma_{\varepsilon 2}}$$
$$\Phi\left(\frac{Z^R_{i,1}}{\sigma_{\eta 1}}\right)^{R^1_i}\left[1-\Phi\left(\frac{Z^R_{i,1}}{\sigma_{\eta 1}}\right)\right]^{1-R^1_i}\Phi\left(\frac{Z^R_{i,2}}{\sigma_{\eta 2}}\right)^{R^2_i}\left[1-\Phi\left(\frac{Z^R_{i,2}}{\sigma_{\eta 2}}\right)\right]^{1-R^2_i}$$
$$\Phi\left(\frac{Z^C_i}{\sigma_{\eta 3}}\right)^{C_i}\left[1-\Phi\left(\frac{Z^C_i}{\sigma_{\eta 3}}\right)\right]^{1-C_i}.$$

Let π_k be the unconditional probability that an individual is type k, then the likelihood function is given by:

$$L(\Omega|GPA,X,G3,W,R,C)=\prod_{i=1}^{N}\left(\sum_{k=1}^{K}\pi_k L_i(\Omega|GPA,X,G3,W,R,C;\tau_i=k)\right). \tag{1.6}$$

The estimated parameters are the Ω and $\{\pi_k\}_{k=1}^K$ that maximize L.[26] As is common in these types of models, the standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Results

In Appendix A.1 we present the point estimates and their standard errors for the two estimation samples. Before performing two exercises to study the effect of grade retention on juvenile crime, there are three aspects of our estimation results worth noting. First, all the signs are as expected and most of the point estimates are statistically significant. Second, the results from these two estimation samples are qualitatively similar. Third, in the case of the restricted sample, the three unobserved types collapse into one, which reinforces the idea that when we focus on the students who are at the margin of grade retention (our preferred specification) there are no unobserved differences among the students below and above the threshold.

We run two simulation exercises to assess the effect of grade retention on crime. The first exercise is to simulate the marginal effect of grade retention in the first period and in the second period on juvenile crime, without dynamic. Thus, we use

[26] We minimize $-L(\Omega|GPA,X,G3,W,R,C)$ using the Matlab solver *fminsearch*.

the point estimates of γ^2 to simulate the effect of grade retention in the first period on juvenile crime, without considering its indirect effect through academic performance and grade retention in the second period. In other words, we only use equation (1.5) in the simulation. In terms of the effect of grade retention in the second period on crime, our model does not have dynamics.

Table 1.6 (panel A) shows the following results from the first exercise. Using the restricted estimation sample we see that grade retention in the first period decreases the probability of crime by 2.1 pp. And that grade retention in the second period increases the probability of crime by 3.8 pp for those students who were subject to a grade retention in the first period and by 4.6 pp for those students who were retained for the first time. For the full estimation sample, first-period grade retention increases the probability of crime by 0.01 pp and second-period grade retention increases the probability of crime by 4.9 pp (for those retained in the first period) and 5.3 pp (for those retained for the first time).[27] These results confirm our concerns about the dynamics in the sense that the negative and strong effect we get from our RD estimation is mainly driven by the difference between the effect of grade retention in the first period versus the second period rather than a relevant negative effect of grade retention on juvenile crime.

To be sure about this interpretation, we run the second exercise which —by considering the dynamics —seeks to replicate our RD estimation by simulating our estimated model. We take the following steps: First, for each student we simulate GPA in the first period, using equation (1.1). Second, we define as marginal those students whose simulated first-period GPA is between $4.95 - 0.16$ and $4.95 + 0.16$. We only keep the sample of simulated marginal students. Third, we randomly assign the treatment of grade retention in the first period to one-half of the simulated marginal students. Fourth, for the treated and control simulated groups, we simulate equations (1.2), (1.4), and (1.5). We do so considering all the dynamics —namely, the simulation of R^1 affects GPA^2, the simulation of R^1 and GPA^2 affect R^2, and the simulation of R^1, GPA^2, and R^2 impact C. Finally, given all these simulations, we can evaluate the (full dynamic) impact of R^1 on GPA^2, R^2, and C.

Table 1.6 (panel B) shows the results from this model-based RD simulation. Using the restricted estimation sample and taking into account all the dynamics of the model, we see that first-period grade retention decreases the probability of juvenile crime by 5 pp, increases second-period GPA by 0.12 points, and decreases second-period grade retention probability by 26.8 pp. In the case of the full estimation

[27] In a previous version of this paper, we estimated the effect of grade retention (at 3rd-5th grades) on juvenile crime, finding an increase of 4.6 pp. To the extent that 3rd-5th grades are close to our second-period definition, it is remarkable how similar are the magnitudes from our simulation with the effects found in the previous paper.

sample, the figures are 1.4 pp, 0.08 points, and 14.6 pp, respectively. Notice that these numbers are qualitatively (and to a lesser degree quantitatively) equivalent to our RD estimations, particularly in the case of the restricted sample. This similarity is reassuring as these marginal effects are moments that we do not directly use in estimating our model. Therefore, the simulations from this second exercise strongly support the interpretation that the results from our RD estimation are not driven by a direct and relevant negative effect of grade retention on juvenile crime, but they are mainly driven by a combination of a negative effect of grade retention in early primary grades on grade retention in later primary grades, with an increasing impact of grade retention on juvenile crime as students progress through primary grades.

Table 1.6: Model Simulations

	Estimation sample	
	Restricted	Full
Panel A: Direct effects		
Grade retention in low grades on crime	-0.021	0.001
Grade retention in high grades on crime (first repetition)	0.046	0.053
Grade retention in high grades on crime (second repetition)	0.038	0.049
Panel B: Direct + indirect effects		
Grade retention in low grades on future academic performance	0.120	0.080
Grade retention in low grades on future grade retention	-0.268	-0.146
Grade retention in low grades on crime	-0.050	-0.014

Notes: This table presents the results from two simulation exercises, considering two estimation samples: Restricted (GPA between $4.95 - 0.16$ and $4.95 + 0.16$, $N = 1,787$) and Full ($N = 11,813$). Panel (A) shows the results from the simulation of the marginal effect of grade retention in low grades and in high grades on crime, without dynamic. Low grades are our first period and high grades are our second period (both in primary school). Panel (B) shows the simulation of the effect of grade retention in the low grades but considering the dynamic, namely, the simulation of R^1 affects GPA^2, the simulation of R^1 and GPA^2 affect R^2, and the simulation of R^1, GPA^2, and R^2 impacts C.

1.6 Conclusion

This is the first paper that estimates a causal effect of grade retention on juvenile crime in a developing country. We implement the standard fuzzy RD approach developed by Calonico, Cattaneo, and Titiunik and Calonico et al. by exploiting a discontinuity in the probability of not being promoted to the next grade that is produced by a grade retention rule in the Chilean educational system. Our results show that repeating a grade —for the first time —in the 2nd or 3rd grade decreases the probability of committing a crime as a juvenile by 14.5 pp and by 10.7 pp for a severe crime.

In addition to this empirical approach, we estimate a semi-structural dynamic model that is crucial to correctly interpreting the RD results in order to guide a policy discussion. More specifically, model simulations show that the decrease in the probability of juvenile crime is not because of a negative and relevant direct effect of grade retention on juvenile crime but is due to the impact of grade retention on future grade retention probability. Hence, our RD results (which show negative and very large effects) are driven by grade retention timing, given that grade retention in the later grades of primary education has a positive and much more relevant effect on crime than the direct effect in early grades. The insights produced by this dynamic model may be very useful for understanding the heterogeneity that we observe in the literature regarding the effect of grade retention on juvenile crime and how this effect depends on the timing of the retention.

The evidence from this paper calls into question the appropriateness of grade retention as a public policy. And this concern becomes even more relevant in the context of Chile, a developing country with high rates of grade retention. If policymakers continue to support this practice, our results indicate that the optimal policy is to retain students in early grades when their performance is around the threshold as a way to decrease the probability of grade retention in late primary school grades.

That said, any interpretation of our findings should consider that we do not take into account other aspects of this policy. Our approach is silent, for example, on how the threat of retention could serve as an incentive for all students to exert more effort (see, for instance, Koppensteiner). Therefore, our results should be considered as only one part of the story and a call for a more comprehensive evaluation of grade retention as a recurrent educational policy.

Chapter 2

The equity and efficiency effects of a relative GPA reward in college admissions

2.1 Introduction

The notion of higher education, especially at selective colleges, as a vehicle for upward social mobility makes the issue of access to these programs policy relevant (Autor; Chetty et al.; Turner). Admission criteria at selective colleges typically rely on a combination of standardized tests and high school grades, but consistent evidence of test score disparities between students from different backgrounds raises concerns about the equity implications of these rules (Rothstein; Card and Rothstein; Zwick and Greif Green). Interventions such as top-percent programs and affirmative action policies are examples of systemic efforts to narrow admission gaps between students from different backgrounds.[1] However, there is wide disagreement on the effects of such interventions on the students they are designed to help, and on other students who are potentially harmed by the introduction of preferences for disadvantaged students.[2]

[1] For California's "Eligibility in the Local Context" see Bleemer, and Black, Denning, and Rothstein for Texas Top Percent Policy. For Brazil's affirmative action Otero, Barahona, and Dobbin and Mello and Bagde, Epple, and Taylor study an affirmative action policy in India.

[2] Dillon and Smith highlight this potential trade-off between equity and efficiency. In the case of California, Arcidiacono and Lovenheim find mixed evidence on the benefit of admission through affirmative action. On the other hand Bleemer presents evidence that supports that the benefit of more selective university enrollment is greater for affirmative action underrepresented minority enrollees. Moreover, Bleemer is one of the first studies to attempt to quantify impacts on the

In this paper I use detailed student records, combined with the admissions formulas used by selective college programs in Chile, to evaluate the equity and efficiency impacts of a 2013 reform designed to improve access to the country's most selective programs for students from disadvantaged high schools. Prior to the reform, students submitted ranked lists of selective college programs to a centralized system using a single offer deferred acceptance (DA) algorithm to rank students and allocate offers of admission.[3] Each college program (e.g., Mechanical Engineering at the University of Chile) used a combination of high school GPA and scores on a standardized test (the "PSU" test) to rank students. The 2013 reform introduced a third component, based on the difference between the student's GPA and the historical mean GPA at her high school. This "GPA^{+}" component was designed to boost the admission chances for students who performed much better than the average for students from the same high school, partly offsetting the lower average PSU scores and lower average GPAs at relatively disadvantaged schools in Chile.

The introduction of this new component in the admissions formulas created three groups of students: (1) those who were admitted to a higher-ranked program under the new formula (a group I call the "pulled-up"), (2) those who lost access to the program they would have been admitted to in the absence of the reform, and were instead admitted to a lower-ranked program (a group I call the "pushed-down"), and (3) those whose admissions outcomes were unaffected. The available data allow me to identify all three groups in the first year of the new system (2013). I am also able to identify the same three groups who would have been present if the reform had been adopted in 2012. I then conduct a simple difference-in-differences (DD) analysis of enrollment, persistence, graduation, and post-graduation outcomes, treating the pulled-up and pushed-down students as separate treated groups and the unaffected students as a control group. In a robustness analysis, I show that the impacts from this DD approach are very similar to the effects implied by a regression discontinuity (RD) approach, focusing on students who narrowly win or lose access to their top-ranked program choices.

I find that, as intended, the pulled-up group included students from lower-income and less-educated families who attended mainly public schools. These students accepted their admissions offers at higher rates than in the previous year and ended up in more selective programs with higher-scoring peers. Over the next 8 years, I find that they graduated from significantly higher-ranked programs than their comparisons from the previous year, though their eventual rate of completing a bachelor's

winners and losers from a top percent plan, using a structural model of admissions for students in the University of California. His findings suggest that the gains for the pulled-up group are larger in magnitude that the losses for the pushed-down group.

[3]The DA is based on Gale and Shapley and described with detail in Rios et al.

degree was nearly identical. Preliminary results for their first few years of labor market entry show, if anything, small increases in earnings. For pushed-down students, the results are largely symmetric (though of slightly smaller magnitude). These students, who tended to come from higher-income and better-educated families, were less likely to accept their admission offers than the comparison group from the previous year, and more likely to skip the year, retake the PSU test, and re-enter the admission pool the next year. They end up graduating with a BA at the same rate as the previous cohort but from less selective programs and colleges outside the selective system. Given the comparability of the impacts on the winners and losers from the reform, the evidence suggests that the GPA^+ boost led to some improvement in the equity of the selective admissions system in Chile, with no change in efficiency.

The next section of the paper begins with an overview of the Chilean college system, which includes both selective institutions (which participate in the centralized admission system) and non-selective institutions (which charge relatively high tuition and use their own admissions rules).[4] At the end of each year, after taking the PSU, students submit a rank order list (ROL) of preferences to the centralized admission system. Programs rank students based on GPA and PSU test scores, with different programs using different weights for the two components.[5] The DA algorithm generates a single admission offer for each applicant: students who are unsatisfied with this choice can choose to take a year out of school, then retake the PSU and reapply the following year, or enroll in a program in the non-selective system. For each cohort of applicants, I have access to their rank order list of programs, and information on the ranking rules used by different programs. Using these data I am able to reproduce the admission offers for 99.9% of the applicants from cohorts before and after the reform. I also observe student enrollment (by college and program) and graduation outcomes for each student, including those in selective and non-selective colleges.

To evaluate the 2013 GPA^+ reform, I use data on the 2013 applicants, and on the numbers of students offered admission to each program, but adjust the ranking rules of each program to take out the GPA^+ component. I then re-run the DA algorithm to generate admissions for the 2013 cohort in the absence of the reform. Comparing admission offers with and without the reform identifies the pulled-up and pushed-down students who win or lose access to a higher-ranked program, as well as the relatively large ($\sim 90\%$) of students whose admissions offers are the same. To measure the causal effects of the reform on the enrollment and graduation outcomes of the pulled-up and pushed down-students, I use the 2012 cohort of applicants and

[4]They use private admission requirements, which limits the knowledge of how students are ranked/selected if excess demand occurs.

[5]Some programs also have additional restrictions as minimum PSU scores and minimum application scores.

compare their actual admission offers to those they would have received if the GPA$^+$ reform had been adopted one year earlier. This identifies potentially pulled-up and pushed-down groups in 2012. Since these groups were not exposed to the reform, but their ranked lists and PSU/GPA/GPA$^+$ performance measures are very similar to those of the same groups in 2013, their outcomes form counterfactuals for the pulled-up and pushed-down groups in 2013 (after adjusting for economy-wide trends using the changes in outcomes of the unaffected groups using a DD approach).

This DD approach allows me to measure the separate effects of the reform on the winners and losers from the reform, and test whether the gains in outcomes for the winners are as large as the losses for the losers. However, its validity rests on two key assumptions. First, I have to assume that in the absence of the reform, the trend in the outcomes of interest for pulled-up, pushed-down, and unaffected groups would have evolved similarly - the so-called "parallel trends" assumption. I test this assumption by comparing the 2011 and 2012 cohorts. Following the same simulation strategy to classify students into the three relevant groups I estimate the same difference-in-differences specification for cohorts for which no reform was implemented (2011 and 2012). I find no significant difference between them when no reform is implemented.

A second key assumption is that the rank order list reported by students doesn't change with the incorporation of GPA$^+$ in the application score. With no restrictions in the report of preferences, the dominant strategy for the DA algorithm is to report preferences truthfully (Gale and Shapley; Roth). While the Chilean system limits students to submitting just 10 choices, most students list fewer than 10, suggesting that most students had no incentive to change their lists in the presence of the GPA$^+$ boost (Haeringer and Klijn; Pathak and Sönmez). Nevertheless, some recent papers suggest that reported ranks, even under a DA system, depend on the probability of admission (Fack, Grenet, and He; Larroucau and Rios). If so, some students who received a relatively large GPA$^+$ boost may have changed their reported list of preferred schools in 2013, relative to what they would have reported in 2012. To check that this behavior is not driving my results I estimate the same models in a sample that excludes students with "very high" boost scores.[6] I find the same qualitative and quantitative results.

As a further validation exercise, I implement a regression discontinuity (RD) design, which does not rely on previous cohort comparisons. Specifically, I begin by estimating my DD models for the (relatively large) subset of pulled-up and pushed-

[6] I compare the admission selectivity of the programs ranked in the first choice for students with the same boost before and after the boost was implemented. Only students with a boost higher than 195 points list on average more selective programs after the reform. I define the checking sample as students with boost scores lower than 150 to be conservative.

down applicants who gain or lose access to their top-ranked program because of the GPA$^+$ reform. The impacts of the reform on this group are very similar to the impacts on the overall groups. I then conduct an RD analysis using a sample of students whose admission scores (under the 2013 rules) are relatively close to the cutoff for their first-ranked choice, and using as a running variable their admission score as determined by that choice.[7] I find that the impacts of passing the threshold for the first-choice program are comparable in sign and magnitude to the DD estimates for the pulled-up and pushed-down groups. The estimates suggest that the winners from the 2013 reform experienced a significant gain in the selectivity of the program to which they were initially offered admission, and of the program from which they eventually graduate (which in most cases is the same), with no effect on BA completion in the 8 years after the application round. Likewise, the losers experienced a significant loss in the selectivity of the program to which they were initially offered admission, and of the program from which they eventually graduate, with again no effect on BA completion.

My results also align with the results from other equity admission interventions that find that access-oriented admission policies at selective universities can promote economic mobility without efficiency losses (Otero, Barahona, and Dobbin; Bleemer; Black, Denning, and Rothstein). Consistently with the results reported in Black, Denning, and Rothstein for the Texas Top Percent policy, I find similar graduation rates (inferred) for pulled-up students than for the average students pre-reform, suggesting that pulled-up students did not struggle more.

This paper contributes to the understanding of equity admission interventions and the effect of admission to more selective universities for students who would not normally have access to them (Black, Denning, and Rothstein; Bleemer; Arcidiacono and Lovenheim; Arcidiacono, Aucejo, and Hotz; Otero, Barahona, and Dobbin; Mello; Bagde, Epple, and Taylor). I build on prior empirical research employing a difference-in-differences approach, and take advantage of the transparency of the admission criteria in order to precisely identify the treatment groups resulting from the admission reform. Unlike earlier studies, this admissions change affected the full spectrum of selective colleges, not just access to a single institution. Thus, I study the effect on the entire population of applicants and on the entire college system (selective and non-selective institutions). Contrary to the mismatch hypothesis (Sowell), which states that low-test students targeted by access-oriented admission programs, like affirmative action, would be better off by attending programs where

[7] Programs have additional restrictions, like minimum test scores, that can make students above the cutoff not being offered admission into their first choice. These restrictions are easier to incorporate into the DD strategy.

they match their peer characteristics, I find that the probability of graduation from the admission to a more selective program does not decrease.[8] This paper also contributes to the early, but growing, literature that evaluates outcomes from changes in the assignment mechanism, in this case, which inputs are used for the assignment score (Agarwal, Hodgson, and Somaini; Otero, Barahona, and Dobbin; Larroucau and Rios).

2.2 Related literature

There is a significant body of literature devoted to studying the returns to college, and more specifically, the returns for varying levels of quality or selectivity. In particular, Dale and Krueger; Black and Smith; Lindahl and Regnér; Dale and Krueger highlight the difficulties of deriving causal estimates from observational data. Recently, numerous studies have used a regression discontinuity strategy to adjust for selection bias and have shown that applicants at admissions thresholds gain from admission into selective institutions (e.g., Hoekstra; Zimmerman; Anelli).[9] In addition, Cohodes and Goodman; Goodman et al.; Zimmerman present evidence from the United States that attending a selective university tends to increase graduation rates. In conclusion, the majority of evidence suggests that college quality has a beneficial impact on student performance, although this result is not universal.

However, these methods may be inadequate for evaluating the effectiveness of access-oriented policies. Students at the margin may differ from those who are targeted by the admission policies. Dale and Krueger provides evidence of heterogeneous returns to selective degrees in the United States - positive for underrepresented groups but zero on average - by analyzing the differences in outcomes for students with similar sets of admission offers but different enrollment decisions. Zimmerman and Hastings, Kane, and Staiger document heterogeneous effect for the case of Chile in terms of field of study and family income.[10] Additionally, treatment effects for those outside of the discontinuity may vary. I expand upon the research that em-

[8] Several papers study the mismatch hypothesis with varying results, see for example Sander and Taylor; Arcidiacono and Lovenheim; Rothstein and Yoon; Bleemer; Arcidiacono et al.

[9] Another body of research focuses on the differential returns to fields of study; see for example Kirkeboen, Leuven, and Mogstad and Hastings, Kane, and Staiger

[10] Zimmerman argues that the greatest returns to top business program attendance in Chile apply only to students from high-income families. Compared to Hastings, Neilson, and Zimmerman, my regression discontinuity analysis provides larger results. This discrepancy is expected since I only examine threshold crossing for the first choice, which results in greater effects than other threshold crossings. Prior studies averaged across all thresholds.

ploys differences-in-differences to examine the consequences beyond the admissions threshold.[11]

My paper contributes to the literature by evaluating the effects of access-oriented policies to selective programs not only evaluating the effects on the targeted group of students but also on the displaced students. In this sense, my paper is most closely connected to Black, Denning, and Rothstein, however, I take advantage of my setting to construct the treatment groups intuitively and transparently. The relationship between selectivity and outcomes for the two affected categories of students is needed to evaluate the efficiency impact on the entire system. It will be beneficial if institutions with a greater level of selectivity have more and better learning materials. In contrast, it can be negative if students increase their likelihood of poor performance and school withdrawal. In the situation of differential impacts, student-resorting policies have the potential to generate both efficient and equitable benefits or costs. In order to evaluate efficiency I consider enrollment and medium-term outcomes such as dropout, college graduation, and earnings in the entire system. My results align with the results from other equity admission interventions that find that access-oriented admission policies at selective universities can promote economic mobility without efficiency losses (Otero, Barahona, and Dobbin; Bleemer; Black, Denning, and Rothstein).

The research on the effects on graduation and earnings of access-oriented policies focuses mostly on affirmative action and Top N percent programs (Arcidiacono et al.; Arcidiacono and Lovenheim; Rothstein and Yoon; Bleemer; Otero, Barahona, and Dobbin; Mello; Bagde, Epple, and Taylor; Black, Denning, and Rothstein; Bleemer; Kapor et al.). The "percent plans" implemented in Florida, California, and Texas ensured admission to the public university systems for students with high grades compared to their high school peers, independent of their standardized test scores. The Chilean reform is similar to these policies in that it increases the likelihood of admission for students with strong grades and is demographically blind. However, there are numerous significant distinctions. The Chilean reform did not ensure access but rather increased the likelihood. Related to this, another distinction is that the Chilean reform compares current students to prior students from the same school, whereas the percent plans compared students from the same cohort. A further advantage of the Chilean context is the transparency of admission rules. The majority of the college admission systems in which access-oriented policies have been studied have some arbitrary component or they are structured in such a way that students could behave strategically to take advantage of changes in the admission policies.

[11] Otero, Barahona, and Dobbin overcomes this challenge with a combination of admission thresholds and an exogenous score shifter.

Cullen, Long, and Reback; Estevan et al.; Mello analyze the school switching behavior for the US and Brazil.Concha-Arriagada shows that this occurred in Chile during the second and third years following the implementation of the reform but the problem was quickly resolved in 2016.

Much of the research on affirmative action has centered on measuring academic mismatch. The mismatch hypothesis posits that graduation rates for minority students who attended selective post-secondary institutions would be lower than for those who attended colleges and universities where their academic credentials are better matched to the institutional average. However, results have not been conclusive (Loury and Garman; Rothstein and Yoon; Sander and Taylor; Dillon and Smith; Arcidiacono et al.; Arcidiacono et al.; Bleemer). My setting is ideal to evaluate this hypothesis. Similar to Bleemer for the case of California, I find that the benefits of more-selective enrollment are at least as large for high-GPA students whose low standardized test scores would have normally disqualified them from selective universities as they are for the higher-standardized test students admitted to those universities and that the graduation rate for the pulled-up students was roughly equivalent to the average for the non-affected students.

A closely connected literature evaluates the mismatch hypothesis in the particular subgroup of STEM programs Arcidiacono, Aucejo, and Hotz; Bleemer. The STEM mismatch hypothesis holds that students admitted through access-oriented policies are less persistent in STEM fields than they would be at universities with fewer admission requirements. Contrary to what previous studies show (Arcidiacono and Lovenheim; Bleemer; Mountjoy and Hickman), the evidence for the Chilean case suggests that students pushed into more selective STEM programs by the reform have a higher probability of graduation. Even though the reform increases the probability of enrolling in a STEM program for pushed-up students, the majority of students applying to STEM programs in the pushed-up group had a fallback option of a STEM program, therefore, they were affected by getting access to better quality programs.

Lastly, this paper also relates to other studies interested in the same admission reform to answer different questions. The most related paper, Larroucau, Ríos, and Mizala evaluates the compositions of the students affected by the reform using the same simulation approach as this paper. Concha-Arriagada also relies on similar simulations to study the strategic behavior of students in 2015, after students learn about the construction of the relative GPA boost and before the policy was fixed to address the strategic behavior. In a similar spirit, Fajnzylber, Lara, and León evaluate the effects of the reform in terms of the GPA inflation and learning effort. Finally, Larroucau and Rios use the variation from 2013 to 2014 in the weights associated with the relative GPA component to estimate models of preferences for program choices.

2.3 Context

The Chilean college admission system is an ideal setting to evaluate the effects of an access-oriented admission intervention like the 2013 reform. The reform introduced a new component based on the student's relative GPA, designed to improve equity in the system. The transparency of the system, together with the availability of rich administrative data allows for the simulation of admission offers with and without the new GPA^{+} component even in years before the reform was implemented, facilitating the construction of meaningful counterfactuals for winners and losers of the reform.

Chilean college admission system

The admission process to selective universities in Chile is a centralized score-based meritocracy, based solely on standardized admission test scores and the high school GPA score of the students. The assignment mechanism - which uses a deferred acceptance (DA) algorithm- generates a seemingly strategy-proof environment and can be replicated when admission preferences, program vacancies, and application scores are available. I discuss in detail these key characteristics of the implementation of my empirical strategy, particularly the identification of the two treatment groups.

The college system and application procedure The Chilean college system has selective (public and private) and non-selective (private) colleges.[12] To enroll in a selective university students have to (i) graduate from high school, (ii) take the standardized admission test at the end of the academic year, and (iii) submit a rank-ordered list of their preferences to the centralized admission system after learning about their test results. This process happens once a year and students can enroll only if they get an admission offer. To enroll in a non-selective college, students have to apply directly and follow the requirements of each institution.[13]

The admission process is organized around programs, instead of majors and universities. Programs have a highly fixed curriculum (which makes switching programs without going again through the application process hard and not common) with expected times for graduation between 4 to 7 years (5 being the mode). In most programs, students earn an academic degree after 4 years but they are required to attend a 5th year and pass a licensing exam to earn their professional degree and

[12] In 2012 and 2013 the selective system was composed of 33 universities, which represented around 60% of college students.

[13] In most of the cases colleges require the admission test score but don't set minimums for admission. Therefore, the restriction is a budgetary constrains.

complete graduation. Programs provide complete certification for most occupations, such as architecture, law, or medicine. This characteristic of the Chilean college system makes the relationship between college and labor market outcomes tighter compared to other settings.

The centralized admission process was established in the late 1960s in combination with an admission test (in the same spirit as the SAT) and a single-offer assignment mechanism based on a student-proposing deferred acceptance (DA) algorithm (Gale and Shapley; Abdulkadiroğlu, Pathak, and Roth). Its development and implementation in the country were led by Erika Grassau.[14] New admission tests were redesigned at the beginning of 2000s and consist of a mandatory math and verbal exam and one additional exam that could be science or history. Tests are taken simultaneously at a national level by the end of the academic year.[15] After scores are published (tests and GPA scores), students can start their application - exclusively online through the Department of Evaluation, Measurement and Educational Registration (DEMRE for its acronym in Spanish) website and without any monetary cost - by submitting a list with no more than ten programs, ranked in strict order of preference (their Rank Order List - ROL).[16] Once the application period is finished, students are assigned to programs with the DA algorithm.

Participation in the admission process is the only channel for students to enroll in any selective program.[17] Because students with higher application scores are more likely to be offered admission to a program than a student with a lower application score, and selection can only be based on that, it is considered a score-based meritocratic system. A program is considered more selective than others if the application score of the last student admitted - the program cutoff score - is higher. The application score is a program-specific index that weights students' high school GPA and standardized test scores.

[14]It is surprising the lack of recognition given to Erika Grassau and her team in charge of implementing that reform, considering how ahead of time it was when compared with the boom of the implementation of DA mechanisms in the last decade.

[15]The Chilean academic year normally goes from March to December, but it is shortened to November in the last high school year

[16]To help applicants in their decision-making, DEMRE distributes a directory that provides an overview of the university admission process, key dates, information about vacancies, extra requirements, and the application score formula for each program for each university. While waiting for their results students can access a simulation mode site with a help video that explicitly states "When selected in one of the preferences all the following ones are eliminated, therefore it is very important the strict order of preferences from higher to lower personal interest."

[17]There are some special admission channels like switching students or students with disabilities but among those quotas admission score is always the selection criteria. This paper focuses on the regular admission channel.

Deferred acceptance algorithm The Deferred Acceptance (DA) algorithm is the assignment procedure used to match students to programs, taking into consideration their preferences and the program vacancies.[18] The algorithm can be described as follows: In the initial step, each student proposes to their most preferred program listed in their ROL. Programs provisionally accept students based on their application scores until they fill their total number of seats, rejecting the rest. In subsequent cycles, rejected students propose to their most-preferred program among those that have not previously rejected them, and programs reject provisionally accepted applicants with lower application scores. This process iterates until all students are assigned to a single program or all unassigned students have been rejected by every program they have ranked. See Rios et al. for a thorough description.

A studied theoretical characteristic of the DA mechanism is that it is strategy-proof, which makes reference to the fact that listing programs in order of true preferences is a weakly dominant strategy when students are allowed to rank every program, i.e. it cannot be manipulated by misrepresenting preferences (Dubins and Freedman; Roth). In the Chilean case, students are constrained to list only 10 choices, with extra conditions for some universities.[19]. Table 2.1 shows that 90% of applicants rank less than 10 programs with a mode of 3, in which case truthful reporting is a dominant strategy (Haeringer and Klijn; Pathak and Sönmez). Assumptions over the rank order list and details about the assignment mechanisms are used to simulate admissions with and without the relative GPA measure. Section 2.4 discuss this procedure.

Relative GPA reform

The relative GPA reform created a grade-based measure that augments the admission criteria with a performance measure that takes account of between-school differences and boosts the admission chances for good students from schools with relatively low standardized test scores. The GPA^+ is based on the grades of a student relative to the historical distribution of GPAs at his or her high school and adds a positive boost to the GPA of students who score above the historical mean, with a maximum boost for those who score above the maximum past score at their school ($GPA^+ = GPA + \text{relative boost}$).

[18]The variant of the student-proposing DA algorithm used by DEMRE establishes that all tied students for the last seat of a program must be admitted.

[19]Universidad de Chile and Pontificia Universidad Catolica de Chile limit the applications to their programs, in order to be valid, to the first 4 preferences. For a detailed analysis of how this could affect the report of preferences see Lafortune, Figueroa, and Saenz

Equity concerns around college admission in the 1960s are what motivated the current admission system (meritocratic and transparent). Around the 2000s the admission test was changed in order to address socioeconomic differences in college admission but the socioeconomic gap in test scores persisted, even after controlling for income and parents' education. This evidence fueled a public debate that highlighted the need for a system able to identify high-ability students even when education conditions for them were not optimal to perform well in standardized test scores.

In the second half of the 2012 academic year, the organization in charge of coordinating selective universities (CRUCH for its acronym in Spanish) informed the incorporation of a third element to calculate students' application scores in the 2013 admission process. The timing was such that students and programs had no scope for strategic responses, as students already have their GPA scores determined and universities have already made their capacity decisions.[20] Before the reform, application score (s_{ij}) for a student i to a program j was calculated as:

$$s_{ij} = \alpha_j \text{Tests Scores}_i + \beta_j \text{GPA}_i$$

The weights α_j and β_j were chosen by the programs under some minimum restrictions defined by the DEMRE such that $\alpha_j + \beta_j = 1$.[21] After the reform was implemented, the GPA^+ measure was included in the formula

$$s'_{ij} = \alpha'_j \text{Tests Scores}_i + \beta'_j \text{GPA}_i + \gamma'_j \text{GPA}^+_i$$

with $\alpha'_j + \beta'_j + \gamma'_j = 1$. For its first year, γ'_j was fixed at a mandatory 10% for all the programs. From Figure 2.1 we can see that most of the programs opted for reducing the weight on β_j to allocate the 10% for the GPA^+ measure, therefore most of the variation observed in allocations comes from the introduction of the relative boost.

The proposed new component was designed to make more competitive the application of students that performed well at their high school by awarding them a boost to their GPA score if they perform above their school average ($\text{GPA}^+ = \text{GPA}$ + relative boost). In Chile, grades are not fully curbed and they have an implicit reference to the minimum content expected by the national curriculum on each subject by year. Due to this, even the best student from a disadvantaged school that struggles to cover the minimum content can have a very low GPA score. The GPA^+ component was designed such that with the boost, students that perform at the top of their school GPA distribution have a GPA^+ score that corresponds to that. By

[20] The literal translation of the reform's name is "Ranking", which is misleading. Given that the score is assigned in relationship with the student's educational context rather than their class ranking, I will refer to it as relative GPA reform rather than Ranking reform.

[21] With a minimum 10% in each of the component.

making the application score of good-performance students higher, the reform helped them access programs that would have rejected them when their application score was lower.

Relative GPA measure in detail The relative GPA (GPA$^+$) measure is based on the GPA score of the student, but it is adjusted with a boost that depends on the historical average ($\overline{GPA}$) and the historical maximum high school GPA of their high school ($\max GPA$). The historical average and the historical maximum are constructed based on the high school GPAs of the students from the previous 3 cohorts at that school. It was chosen as a reference for the within-school measure to avoid within-classmates' competition. The formula to calculate the (GPA$^+$) score is the following

$$GPA_i^+ = \begin{cases} GPA_i & \text{if } GPA_i < \overline{GPA} \\ \overline{GPA} + \frac{850}{\max GPA}(GPA_i - \overline{GPA}) & \text{if } GPA_i \in [\overline{GPA}, \max GPA] \\ 850 & \text{if } GPA_i > \max GPA \end{cases}$$

Students with a GPA equal to or lower than the historical average at their schools have a relative GPA score equal to their GPA score. Students with a GPA bigger than the historical average but smaller than the historical maximum get their GPA score plus a boost that is determined by the slope of the line that connects the historical average GPA score with the historical maximum, which is for all schools the maximum possible score, 850.[22] This implies that students in this range, from a school with a more spread out high school GPA distribution, will have a smaller boost in terms of score points for each extra point in their GPA. Finally, students that perform above the historical maximum at their high school get the maximum possible score (850), even if the GPA is, measured in application points, very low.

In order to simulate the admission assignment under the new mechanisms defined by the inclusion of the GPA$^+$ for cohorts previous to the implementation of the reform I construct the GPA$^+$ measure for the cohorts 2009 to 2012. According to the reform, students who graduate from cohorts before 2009 or students who didn't attend a school had the relative GPA score equal to their GPA score.

[22]Figure 2.2 correspond to an example to represent the relationship between GPA, GPA$^+$ and the boost.

Data

I focus my analysis on the entire universe of applicants to selective universities during the years 2012 (pre-reform) and 2013 (post-reform). For the first part of the empirical analysis, I construct a unique dataset that replicates college admission offers with and without the inclusion of the relative GPA measure in the admission process for the students in these two cohorts. This allows me to classify students into one of the three possible groups of analysis: pulled-up, pushed-down, or unaffected. To assess human capital acquisition, I add data on annual enrollment and graduation from selected and non-selective colleges for all the applicants to the 2012 and 2013 process. Finally, I add to the analysis information on employment and earnings on the private labor market up to 10 years following their application.

Admission process The relative GPA reform was implemented in the admission process of 2013. For that reason, my analysis focuses on the short and medium-long-term outcomes of all the students that participated in the admission process that year and the year before (2012). I use information from students in the 2011 cohort to validate my research design.[23]

Administrative data at the student level from the admission process was shared upon request by DEMRE. It consists of socioeconomic and demographic information of applicants (gender, date of birth, self-reported family income, and parents' education), applications scores (tests scores, GPA, and relative GPA score), high school characteristics, application information (rank order list of program preferences listed in the application with their final status: valid/invalid, offer/no offer and waitlist), and enrollment information (program, application score, and ranking of preference). This information is mainly used to simulate students' admission under a mechanism that uses two (test scores and GPA) or three (test scores, GPA, and GPA^+) inputs to calculate the application score.

The "new" mechanism incorporates the relative GPA measure (GPA^+) into the application score formula. To compute the relative GPA measure for cohorts before the reform I use information from the national school records on high school performance for the entire population of high schoolers between 2002 and 2011 which is available online at the data platform of the Department of Education.[24] I compute

[23]Even though information for later cohorts is available I don't consider it in my analysis because my empirical strategy is sensitive to the strategic behavior observed during those years. After 2013, some students switched schools in their last year of high school to improve their GPA^+ measurement. This potential for policy manipulation was fixed in the 2015 process.

[24]`https://datosabiertos.mineduc.cl`

the historical average and the historical maximum GPA at each school for each graduation cohort, and then the relative GPA score for students who graduated between 2008 to 2012 in the 2011 and 2012 admission process.[25] Figure B.1 shows a binscatter graph with the boost score - i.e. the extra score relative to GPA- of the relative GPA score for students in application cohorts 2011 to 2013. The x-axis is the GPA score of the student minus the historical average high school GPA at the school of the student, therefore on the positive numbers we see the boost score in application points. Note that 2013 data is directly reported by DEMRE and 2011 and 2012 were calculated using the relative GPA score formula.

I also constructed a dataset with program characteristics like application score weights, application score restrictions, and the total number of seats from the public newsletter with the official information. Application score weights are required to calculate the application score under the two regimes. For each program, application scores under the status quo regime (s_{ij}) are calculated using weights from the 2012 process, and application scores under the GPA$^+$ regime (s'_{ij}) are calculated with 2013 weights.[26]

Enrollment and graduation outcomes To measure the effect of the reform on educational outcomes I track all the students that participate in the application processes of 2012 and 2013 using yearly information on enrollment and graduation provided publicly by the Department of Education. From the admission data, I can observe who got an admission offer and to which program. I create variables to indicate if a student enrolls in their admission offer or if they enroll in a non-selective college instead. By using the enrollment file in the second year ($t = 2$) I check if the student persisted in their admission offer, if they re-apply or switched to a different selective program, if they switched or persisted in a non-selective college, or if they dropped out of college.

Additionally, for each application cohort, I track graduation by the 6th, 7th, and 8th years after application because yearly graduation files were available only up to 2020. I construct 3 graduation measures: (1) program graduation or graduation from their initial admission offer in 2012 or 2013, (2) graduation from some selective university to take into account that students that don't get their desired admission may switch or re-apply for the following years, and (3) graduation from a non-selective college which is always an alternative. Having access to data on the entire

[25]Students can participate in the admission process as many times as they want. The proportion of freshmen and older applicants is around 60% to 40% in each cohort.

[26]Music, arts, and acting programs require an additional aptitude test, which score is not reported separately in the data. For those cases, the application score used for the alternative regime was the same as the one reported originally.

system allows me to measure the complete impact of the reform in the selective system - the one that DEMRE attempts to coordinate-, as well as the impact on the entire college system.

Labor market outcomes To study the effect on earnings of giving access to better programs to students that normally couldn't access them I use information from the Unemployment Insurance (UI) data. The UI data has information on all the dependent workers over 18 years old that participate in the private sector.[27] All the information is aggregated at the treatment group level. For pulled-up, pushed-down, and unaffected students I observe the fraction that was present in the labor market (participation) and bins for their monthly taxable income from 8 to 10 years after the admission process.

2.4 Empirical strategy

The empirical strategy is divided into two parts. First, I simulate the admission mechanisms with and without the relative GPA measure. I classify students into 3 groups based on the admissions simulations: (i) pulled-up, students who gain access to more selective admissions when the third component is considered in the assignment mechanism, (ii) pushed-down, students who lose access to more selective programs with the new mechanism, and (iii) unaffected, students whose admission options are unaffected by the change in the mechanism. By simulating the admissions under the two mechanisms in earlier years, before the reform was implemented, I can identify the groups who would have been pulled-up and pushed-down in those years. This facilitates a difference-in-differences design to estimate the impact of the inclusion of the relative GPA on enrollment, graduation, and earnings for the students affected by the reform.

Identification of treatment groups: pulled-up, pushed-down, and unaffected

The inclusion of the relative GPA measure into the admission process enhanced the equity of the college admission system. Students with relatively low test scores but high GPAs from low-educated and low-income families got admissions into more

[27]Data excludes: (i) workers subject to an apprenticeship contract; (ii) workers under 18 years of age; (iii) private home workers (until October 2020); (iv) pensioners; (v) independent or self-employed workers; and (vi) public sector workers. In a future version of the research, I will be able to include information on public sector workers and person-level data.

selective programs when the third component (GPA$^+$) was considered. There is also a higher representation of females in the pulled-up group of students. Pushed-down students tend to be in higher proportions from private schools, males, and from highly educated and high-income families. Looking at the impact in admission offers induced by the reform, most students affected had an admission one preference up or down with respect to the status-quo regime, they are moved into or out of their 1st preference, and they get a new admission in the same field.

Simulation of the admission mechanism The relative GPA reform impacted the way that students were matched to the programs that they apply. Before its implementation, the application score for a student i applying to a program j was calculated using only 2 inputs: admission test scores e_i and GPA score g_i. With the implementation of the reform, the new application score was calculated based on $s'_{ij}(e_i, g_i, c_i)$. Denote $\mu(\cdot)$ as the matching function defined by the mechanism that uses a Deferred Acceptance algorithm, the information from the pool of applicants, the application scores defined by the programs, and the capacity restrictions of the program. The change in the inputs used by programs to evaluate students defines a new mechanism $\mu'(\cdot)$.

A student i can be characterized by $\theta_i(\succ_i, e_i, g_i, c_i)$ composed of their rank order list $(\succ_i)$ and their scores. In each application year, for some students the admission assignment under both mechanisms will differ, $\mu(\theta_i) \neq \mu'(\theta_i)$, and for others it won't $\mu(\theta_i) = \mu'(\theta_i)$. I classify the pool of applicants into 3 mutually exclusive groups:

- Pulled-Up: $PU_i = 1\{\mu(\theta_i) \prec \mu'(\theta_i)\}$ students who get access to a program ranked higher in their list with the new mechanism μ' than with the old mechanism μ.
- Pushed-Down: $PD_i = 1\{\mu(\theta_i) \succ \mu'(\theta_i)\}$ students who get access to a program ranked lower in their list with the new μ' than with the old mechanism μ.
- Unaffected: $C_i = 1\{\mu(\theta_i) = \mu'(\theta_i)\}$ corresponding to students with access to the same programs with and without the inclusion of the GPA$^+$ measure.

Implementation of admission simulations For each student, in each application process, I start by computing their alternative application score. For students pre-reform this also includes computing the GPA$^+$ score. For each program that the student listed, I use the weights from 2012 and 2013 to calculate the alternative application score (for students in the 2012 cohort I calculate s'_{ij} and for students in 2013 I compute s_{ij}).

I replicate the DA algorithm to simulate the admission assignment of students with the GPA$^+$ measure for pre-reform students ($\hat{\mu'}(\theta_i)$), and without it for post-reform students ($\hat{\mu}(\theta_i)$). In order to test the quality of the replication I simulated the admission assignments using s'_{ij} for cohort 2013; I replicate 99.9% of the real assignment offers.

For each student in the application cohort 2012 or 2013, I compare the simulated admission with the real admission offer and I classify them into the pulled-up (pushed-down) group if the admission assignment with the GPA^+ measure was higher (lower) in the list than the assignment without it. Students are classified as unaffected if the admission program under both regimes is the same.

Simulation assumptions There are three main assumptions needed for the simulation to be valid as a counterfactual under the alternative mechanism.

Assumption 1 *The rank order list of preferences that the students submit would have been the same with and without the reform*

Assumption 1 has two components, one that refers to the stability of preference and one that refers to the reporting behavior. I assume that preferences are stable with respect to the reform, which means that the indirect utility associated with each program does not depend on the components and weights used by the programs to evaluate applicants.

In terms of reporting behavior, I use the traditional approach taken by the literature that establishes that without restrictions on the number of applications, the dominant strategy with a Deferred Acceptance (DA) algorithm is truthful reporting (Gale and Shapley; Dubins and Freedman; Roth). As most centralized admission system, the Chilean application system restrict the application list (up to 10 options), however, because more than 90% of the students list fewer than 10 options, the restrictions can be interpreted as not binding (Haeringer and Klijn; Abdulkadiroğlu and Sönmez; Abdulkadiroğlu, Pathak, Schellenberg, and Walters).

One possible concern rise from the recent literature on mechanisms design and their interest in using the information from the centralized admission systems to estimate school choice demands models (Agarwal, Hodgson, and Somaini; Fack, Grenet, and He; Larroucau and Rios). One way of rationalizing the fact that students don't fill up their application options relates to the idea that reporting behavior is based on students' feasible options. This behavior may violate assumption 1 if students that observe the boost (that potentially could increase the set of desirable options that they will be eligible for) reacted by adding more selective programs to the top of their list. This would create a problem in the identification of the treatment group if students get admitted to the new added programs but similar students that didn't observe the boost (cohort of 2012) didn't get admitted under the simulation (because they didn't list the new options).

To assess this potential threat I first compare the number of admission options listed in 2012 and 2013 by students with a boost (by adding a program to the top of the list, the total number could increase). Students that observed the boost in 2013 are not more likely to have longer application lists than students with the same calculated boost but who didn't observe it (cohort of 2012). Additionally, I check the selectivity of the most preferred program or top-ranked program of students with a boost, in 2012 and 2013. Figure 2.3 show that the selectivity of the first option (measured as the application score of the last person admitted in that program) increased in 2013 only in the highest values of boost score distribution. In order to check for the sensitivity of the results I estimate the results without students with more than 150 points in their boost score (2% of the total sample and a conservative range compare to what is observed in the graph). As discussed in Section 2.9, results don't change qualitatively or quantitatively with this sample restriction.

Assumption 2 *The number of available seats per program each year would have been the same with or without the reform*

Assumption 3 *Standardized test scores and GPA scores would have been the same with and without the reform*

Assumptions 2 and 3 are justified by the fact that the reform was announced in the last half of the academic year. At that point, universities have already made their capacity decisions, and students' average GPA from the 4 years of high school was already determined, therefore there was no scope for strategic responses.[28]

Characterization of treatment groups Table 2.2 shows the characteristics of the group of students identified as pulled-up, pushed-down, and unaffected for cohorts of applicants in 2012 and 2013. Each year, pulled-up and pushed-down applicants account for approximately 4% of the applicant pool. From Table 2.2 we can see that the reform was able to impact the students that were targeted by it. Students in the pulled-up group have better GPAs than those in the unaffected and pushed-down groups; yet, their exam scores are comparable to those in the unaffected group. Looking at pushed-down students, they have low GPAs and high test scores. Moreover, pulled-up students are 3 times less likely to attend a private high school than a

[28] After the first year, there is some evidence, at least anecdotal, about students switching schools in their last year in order to graduate from schools with very low maximum historical GPA in order to gain the maximum score from the GPA^+ component. In 2015 this problem was addressed with a change in the policy, which established that the score was calculated relative to the GPA of the student and the school that they attended each year.

pushed-down student, and looking at family characteristics, pulled-up students come from families with an average income 30% lower than pulled-up students, and their parents are less educated.

Impact of the reform on admission offers Figure 2.4 presents the distribution of pulled-up and pushed-down students based on the number of positions moved in their rankings between the admission assignment with and without GPA$^+$. If the most preferred program that the student could reach without the GPA$^+$ measure was choice 3, but with the inclusion of the boost the student could get into their most preferred option (pulled-up students), then the student was moved 2 positions due to the reform. Figure 2.4 shows that the change in terms of preferences is similar for pulled-up and pushed-down groups and that most of the students affected by the reform were moved one position in the preference list.

A more detailed analysis of the distribution of rankings for admission is presented in Table **??**. Each row presents the number of students with admission assignments in that ranking when the relative GPA measure is considered. Each column presents the total number of students with admission assignments in that preference choice when the GPA$^+$ measure is not considered. Students assigned to the same program in both regimes are classified as unaffected and are presented in the table without background color (table diagonal). The percentage value in each cell corresponds to the proportion of students in that group in that specific ranking combination. The main margins of treatment – changes in the ranking of the admission with and without the reform – correspond to movements into and out of students' 1st preference. The high percentage of students moved between no admission and 1st choice is not explained by a higher proportion of students with short rank order lists but rather due to the bigger proportion of students at the margin of the minimum requirements of not very demanded programs. More specifically, certain programs establish complementary restrictions to admission, such as minimum application scores (taking all the components into consideration) or minimum test score averages. Students in this margin have twice a higher proportion of their total rank order list as invalid due to these extra restrictions.

Finally, Tables 2.4 and 2.3 present the number of pulled-up and pushed-down students in each field with and without the inclusion of the GPA$^+$ component, based on the fields of the admission and simulated admission. For both groups, in most of the cases, students move along their ranking but they stay in the same field (diagonal of the table).

Difference-in-differences design

I estimate the effect of the reform on human capital acquisition and earnings, on the group of pulled-up and pushed-down students. My difference-in-differences design compares the outcomes of students who apply in cohorts after the implementation of the reform - therefore affected by it - versus those in cohorts before the implementation of the reform. With the estimation of the effect of the reform on pulled-up and pushed-down students, I analyze the (outcome) efficiency impact of the reform on the system.

The parameters of interest to evaluate the effect of the inclusion of the relative GPA measure in the admission process can be expressed as the conditional average treatment effect for the group of students pulled-up and pushed-down.

$$\tau(PU) = \mathbb{E}[Y_i(\mu') - Y_i(\mu)|PU_i = 1]$$
$$\tau(PD) = \mathbb{E}[Y_i(\mu') - Y_i(\mu)|PD_i = 1]$$

In the potential outcome framework $Y_i = D_iY_i(1)+(1-D_i)\cdot Y_i(0)$ is the outcome of a student i, and $D_i = 1\{\text{when the relative GPA is used for admission assignment}\}$. The observed outcomes is represented by $Y_i = 1\{t(i) = 2012\}\cdot Y_i(0)+1\{t(i) = 2013\}\cdot Y_i(1)$. Assuming additive separability to capture any changes in time uncorrelated to the determinants of the outcomes with and without the inclusion of the GPA$^+$ measure, I estimate models of the form:

$$Y_i = \beta_1 PU_i + \beta_2 PD_i + \beta_3(PU_i \cdot Post_i) + \beta_4(PD_i \cdot Post_i) + \beta_5 Post_i + X_i'\Gamma + \varepsilon_i$$

where Y_i is the outcome variable of interest to evaluate the reform: enrollment, graduation, and earnings. PU_i indicates if the student belongs to the pulled-up group, PD_i indicates if the student belongs to the pushed-down group, $Post_i$ is an indicator that takes the value of 1 if the students apply post-reform. The omitted group is students that get access to the same programs under both regimes. X_i' is a vector of individual characteristics such as gender, family income, type of school, GPA, and standardized test scores to control for possible changes in the composition characteristics of pulled-up and pushed-down students between 2012 and 2013.[29]

Here β_3 and β_4 are the estimates of the parameter of interest to evaluate the reform. β_3 captures the effect on outcome Y_i of gaining access to the more preferred, but also more selective program due to the inclusion of the GPA$^+$ measure in the

[29]Results are presented with and without controls. Most of the results are quantitatively and statistically unchanged.

admission process. Likewise, β_4 captures the effect of losing access to more selective programs with the reform.[30]

Identification assumption The key identification assumption is that the outcomes for these three groups of students would have evolved similarly for the cohorts 2012 and 2013 if the reform would have not been implemented. I cannot directly test that, however, I conduct a placebo exercise with data from the 2011 application cohort that present suggestive evidence in support of it.

Following the same procedure used for cohort 2012, I start by computing the boost score for each student in 2011, and application scores for each program in their rank order list. With that, and keeping constant the vacancies observed that year I re-run the DA algorithm using the three components application score. Using the simulated admission assignment I classify 2011 students into pulled-up, pushed-down and unaffected. Finally, I estimate the diff-in-diff specification but with the variable $Post_i$ indicating if the student was observed in the 2012 admission process.

Table 2.5 shows the estimates for this placebo exercise, which can be interpreted as the effect on enrollment and graduation for pulled-up and pushed-down students when no reform is implemented. As expected, there is no significant effect suggesting that when no reform is implemented these groups follow a similar trend. The estimates would be biased if the coefficients of interest reflect sample selection resulting from the impact of the reform on the composition of applicants. However, there is no change in the trend of total applicants, and no change in the probability of pulled-up students reapplying compared with the 2011 cohort. There also would be bias in the estimates if there were unexpected changes in 2013 in other determinants of outcomes that differentially affected the three groups. I am aware of no such change.

Notably, the intervention considered for this diff-in-diff evaluation occurred just once, so considerations regarding the calendar time of the comparison group observations, such as those stated by Goodman-Bacon; Baker, Larcker, and Wang; De Chaisemartin and d'Haultfoeuille, do not apply in this context.

2.5 Enrollment results

The change in the admission mechanism due to the inclusion of the relative GPA measure had a large impact on initial enrollment for pulled-up and pushed-down students. However, this change fades out with time; 3 years after the implementation

[30]The new admission program is more preferred by definition of the treatment group, but it has to be more selective because if it wasn't the case, that program would have been reached in the status quo scenario.

of the reform the changes in the probability of enrollment is zero for pulled-up and pushed-down groups.

The difference-in-differences estimates in Table 2.6 show that, for pulled-up students, there is a large effect on the probability of students choosing to enroll in their admission offer. After the reform, pulled-up students are 22 p.p. more likely to enroll in the selective program if they were admitted. This is a 40% effect on enrollment.[31] For pushed-down students the probability of enrollment decreases by 16.7 p.p. The difference (in absolute value) between the effect on enrollment for pulled-up and pushed-down students is significant, indicating that the inclusion of the GPA$^+$ measure improved the system in terms of identifying successful applicants, i.e., there is an increase in the total number of students that decide to enroll once admission is offered.

The total effect on enrollment uncovers changes at two margins: the extensive margin - students that gain or lose the possibility of admission in the selective system - and the intensive margin - students that improve (worsen) their admission in the selective system, but that with or without the reform would have had some admission on the system. On the extensive margin, the reform changed the probability of a student getting access to some selective program in pulled-up and pushed-down students by approximately 20%.

The total effect on initial enrollment is not fully driven by students at the extensive margin. To study the intensive margin, I restrict the sample to students that would have got some admission under the two regimes. Observing the admission offers under the two regimes allows me to correct for the potential selection bias of only observing enrollment if a student actually gets an offer.[32] Columns 5 and 6 of Table 2.6 presents the results restricted to the group of students at the intensive margin. The estimates on initial enrollment for pulled-up students after the reform is smaller (17 p.p.) but still large. Compared with the pushed-down students (11 p.p.), I find evidence of higher intensity of preferences for pulled-up students, i.e., that the reaction, in terms of enrollment decision, from getting access to a program higher in the rank order list is stronger than the reaction from losing access to it, for the pushed-down group.

I summarize the changes in the programs that students attend using traditional measures of quality like selectivity and graduation rate. Table 2.7 shows how the characteristics of the peers and programs that students attend before and after the reform changed. Columns 1 and 2 show the diff-in-diff estimates of a regression in

---

[31] Table B.1 presents the average enrollment rates in the selective system for the 3 groups.

[32] All students in the pulled-up group got an admission offer in 2013 (if not they could not be better than without the GPA$^+$ measure), but not all pulled-up students got an admission offer in 2012 because the reform was still not implemented.

which the dependent variable in one of these average program characteristics before the reform. The first 3 rows show that pulled-up students attend more selective programs after the reform, in the sense that the average student at the program they enroll in had higher test scores and GPAs than the average student at the programs they enroll in before the reform was implemented. Graduation on time is an indicator of the probability that a student graduates in the number of years set by the program; after the reform pulled-up students enroll in programs where the average student is more likely to graduate on time. The results are symmetrical for pushed-down students.[33]

Enrollment up to 4 years after the reform If students are unsatisfied with their initial admission offer students can enroll in a non-selective college or re-apply to the selective system the following year (normally after taking extra test preparation courses). Columns 3 and 4 in Table 2.6 show that in the first year, pushed-down students compensate for the decrease in the probability of enrolling in a selective program by enrolling in the non-selective system. However, the 3.9 p.p. increase in the probability of enrollment in the non-selective system does not offset completely the decrease in the probability of enrollment in the selective system. This means that the reform leads to some pushed-down students not enrolling in any university in the first year after high school.

Table 2.8 shows that pushed-down students are 7 p.p. more likely to reapply to the selective system after the reform was implemented. Table 2.9 presents the changes in enrollment at any program for the pulled-up and pushed-down group up to 4 years after the implementation of the reform using the same diff-in-diff specification. The initial difference in enrollment (even considering non-selective programs) generated by the reform is fully reversed in the second year for pushed-down students. Column 3 shows that 3 years after the implementation of the reform pulled-up students are still 1.5 p.p. more likely to be enrolled relative to before the implementation of the reform. This difference is fully offset 4 years after the reform.

The selectivity of the programs that students attend changed after the reform and a difference persisted throughout time. Table 2.10 presents the change in the average test score of the peers in the same program, up to 4 years after the implementation of the reform. The initial enrollment for pulled-up students is at significantly more selective programs after the reform, however, students in the 2012 cohort seem to react the second year after the implementation of the reform by reapplying and enrolling at more selective programs (which makes the difference in the selectivity

[33]The expected graduation time of the programs that pushed-down students to enroll after the reform are on average 0.07 years shorter.

between the two cohorts to go down). From columns 2, 3, and 4 we see that pulled-up students ended up enrolled in programs with peers with on average 0.1 s.d. higher test scores after the reform was implemented. The behavioral response is similar for pushed-down students. After the reform, students are less likely to enroll and they enroll in less selective programs. However, when they reapply they are able to reach more selective programs (especially 2 years after the reform). Four years after the implementation of the reform there are no differences in the probability of enrollment for push-down students, however, the selectivity of the programs is lower than without the reform (peers have on average 0.08 s.d. smaller test scores).

2.6 Graduation effects

I find that pulled-up students are 8.4 p.p. more likely to complete their initial admission program; pushed-down students have a comparable opposite effect (-8.2 p.p.). An alternate exercise designed to test for the mismatch hypothesis confirms this preliminary evidence against it. Pulled-up and pushed-down students have no effect on the probability of college graduation when considering graduation from any program (and not just from the new programs granted admission as a result of the reform) and the probability for them to remain enrolled due to the delayed enrollment.

Admission program completion There is a positive effect in the probability of completing their initial admission program for pulled-up students, with a comparable opposite effect for pushed-down students. Columns 1-3 of Table 2.11 present the results for graduation from the admission program at different points in time. Consistently, there is a large positive effect (8.4 p.p. increase by 8 years after the reform) of 36% on the likelihood of graduation from the admission program for pulled-up students. Column 4 also shows that pulled-up students are more likely to graduate on time after the implementation of the reform. For pushed-down students the effects on graduation are similar in magnitude but with the opposite sign.

In essence, the reform enabled pulled-up students access to more selective programs which increased their likelihood of enrolling in and graduating from those programs. Putting the graduation effect for pulled-up students into perspective, the implied graduation rate for the marginal student admitted by the relative GPA is 38% (8.4/21.9). This does not differ much from the average graduation rate of unaffected students post-reform (40%) or from the pre-reform level of 39% percent. In addition, the impacts are qualitatively comparable to the findings of other equitable college admission programs, such as Black, Denning, and Rothstein and Bleemer.

Mismatch hypothesis The mismatch hypothesis establishes that applicants with lower test scores targeted by equitable admission policies would benefit from enrolling in less selective universities, where their academic qualifications more closely "match" those of their peers (Sowell). This hypothesis found empirical support on some of the mixed results from the research around affirmative action policies like Arcidiacono and Lovenheim. However, the evidence presented so far for the relative GPA reform contradicts this hypothesis; I interpret the fact that students in the pulled-up group enroll in more selective programs after the reform and increase their probability of graduation from those programs as evidence against the mismatch hypothesis.

Because the main specification doesn't control for the tuple (specific pair of admission programs with and without the GPA$^+$ measure) of admission programs, one possible concern refers to the potential imbalances in the programs that students get admitted with and without the reform, between 2012 and 2013.[34] In order to control for that, I estimate an alternative specification that includes as a control the admission assignment without the reform. This way, I can ensure that all the variation captured by the diff-in-diff comes from pulled-up students with the same admission assignment without the reform and with admission to more selective programs after the reform.[35] Table 2.13 shows the result from this exercise. Contrary to the mismatch hypothesis, more selective admission increased the graduation probability for pulled-up students, with a similar effect than the estimated before (9 p.p.).

STEM In recent years there has been a special interest in STEM degrees, and the focus on this topic for access-oriented policies have not been the exception (Loury and Garman; Holzer and Neumark; Arcidiacono, Aucejo, and Hotz). Arcidiacono, Aucejo, and Hotz study major degrees for the case of California campuses when affirmative action policies were in place; their research states that a better matching of science students to universities by preparation level could increase minority science graduation.

I find that the effect on degree completion in STEM for STEM applicants is positive and significant (6 p.p.). Column 1 in Table 2.14 shows that the relative GPA reform increases the probability for pulled-up students to get admitted in a STEM program. Column 2 presents the effect of enrollment in a STEM program, conditional on students listing some STEM programs in their application, and column 3 also presents enrollment results but focuses on students at the intensive margin of

[34] Students in the pulled-up group are by definition admitted to more selective programs post-reform, but this is relative to their own assignment.

[35] Remember that the definition on pulled-up group is based on the ranking of the preference, but if something was ranked higher and was less selective than the admission assignment without the GPA$^+$ measure, then the algorithm would have assigned the student to that program pre-reform.

treatment. The effect on enrollment (16.9 p.p.) compared to the effect on graduation in the same sample (6.1 p.p.) suggest that the implied graduation rate for the marginal student admitted by the reform is higher than the graduation rate in STEM degrees for the unaffected students in the entire system (36% vs 24%).

College completion There is no effect of the reform (pulled-up or pushed-down group) on human capital acquisition when it is measured as college completion and when the possibility for students to be still enrolled 8 years after the application process is considered. However, pulled-up students are more likely to earn degrees from selective programs after the reform.

Table B.3 shows the average graduation from any program by 6, 7, and 8 years after the implementation of the reform by treatment groups. Notice that graduation from any program captures some of the indirect effects of the reform in reapplications (therefore late enrollment in the selective system) and enrollment in the non-selective system. This could be one of the reasons why, even 8 years after the application process, there are still important changes in graduation rates relative to the previous year, suggesting that the lack of more graduation data limits the full analysis of the reform.

The difference-in-differences estimates for the effect of graduation from any program are presented in Table 2.15. There is no change in college completion for pulled-up and pushed-up students by 7 year after application due to the reform. However, there is a negative effect on graduation 8 years after the reform for pushed-down students, i.e., without the reform they are more likely to have completed some program. I interpret that results as a consequence of the behavioral response in enrollment for pushed-down students. As a consequence of their late enrollment after the reform (they are weaker candidates due to the introduction of the relative GPA and they take more attempts to enroll in the programs that they like) they are more likely to graduate late (even after 8 years from the implementation of the reform). Column 4 of Table 2.15 presents the result when the dependant variable indicates if the student graduate or is still enrolled 8 years after application. The null effect implies that pushed-down students are not acquiring less human capital after the reform.

Table 2.16 presents the results divided by graduation from any selective program and Table 2.17 from any non-selective program. These results also suggest that changes in graduation at 8 years after application for pushed-down students are driven mostly by changes from selective enrollment, which requires a late enrollment if the student wants to enroll in a different program than the admission offered by the new mechanism after the inclusion of the relative GPA measure.

In summary, the reform made pulled-up students more likely to graduate from more selective programs, with no impact on college completion. For pushed-down students, the inclusion of the GPA$^+$ made them less likely to graduate by 8 years after, however, this is not due to a decrease in the probability of college completion but due to a delayed enrollment in selective programs, for some of the students that didn't enroll or didn't stay in the program admitted after the reform.

2.7 Heterogeneity analysis

I first examine the effects of the reform dividing the group of pulled-up and pushed-down students into two groups based on changes of selectivity (measured as the average of test scores) between the admission program with GPA$^+$ and without GPA$^+$. Table 2.18 shows in column 1 that the effects on initial enrollment are positive and larger for pulled-up students with smaller changes in selectivity relative to the group with larger changes in selectivity. Column 2 presents the effect on graduation from the initial admission; the effects are positive for pulled-up students and larger for students with a bigger change in selectivity. For graduation from any program, small increases in selectivity have a detrimental effect on students, but this effect appears to be driven by students taking longer than eight years from their participation in the admissions process to graduate. Results for pushed-down students follow a similar pattern across all the outcomes, students with a bigger reduction in selectivity are less likely to enroll, graduate from the initial admission program, and graduate 8 years from their participation in the admission process from any program; however, the effects are non-significant when the outcome of graduation or still enrolled is considered.

From Table **??** we observe that the main margin in which the reform affected students was increasing (decreasing) the admission of pulled-up (pushed-down) students into their 1st choice. Table 2.19 presents in columns 1 and 2 the main results for this sample, i.e., students moved to and from their 1st choice when the relative GPA was considered. In this sample the effects of the initial admission program are bigger than in the entire population; however, when behavioral responses are considered there is no effect on human capital acquisition 8 years after the implementation of the reform.

Appendix B.1 examines the differential effects by gender, income, and boost score of students on the main outcomes of enrollment, graduation from admission, college completion from any program, and graduation from a selective program. Table B.4 shows the differential effect of enrollment (15 p.p.) only for students with a higher boost. In terms of graduation from admission Table B.5 suggests that the main effect for pulled-up students is driven by the effect on females and students with high boost. There are no differential effects on college completion when any program is considered. I find some indications of variation of impacts across gender, family income, and boost score, but the overall picture is pretty consistent.

2.8 Alternative empirical approach

Using a regression discontinuity (RD) design that permits a direct test of the identification assumptions, and does not rely on previous cohorts, I evaluate the impact of getting access to the most desired program after the implementation of the reform.

I use a regression discontinuity design to estimate the effect on enrollment and graduation of threshold crossing the 1st preference's cutoff because, as shown in Table **??**, the main margin of treatment of the reform is with respect to individuals moving to and from their first preference. I estimate the effects of crossing the admission cutoff for the most preferred program (δ) on enrollment, selectivity of the enrolled program, graduation from the admission program, and any graduation after 8 years using a standard regression discontinuity specification of the form

$$Y_i = f(r_i) + \delta C_i + \eta_i$$

where Y_i is one of the outcomes listed above for individual i; r_i is the difference between the admissions score assigned to i's most preferred program and the admission cutoff score to that program or running variable; $f(r_i)$ is a smooth function (results presented in Appendix B.1 for polynomials of degree 1 to 5) of the running variable (which can change on either side of the cutoff); C_i indicates if i's application score is greater than the cutoff score (so i is admitted to the most preferred choice), and η_i is an error term. I estimate this equation using data from all the programs with excess demand (for which the cutoff is meaningful) on the whole range, with the exception of the linear specification, for which I limit the data to a small score window close to the cutoff.

Table 2.20 provides a summary of the principal results from the RD estimator employing a polynomial of order 3 and the diff-in-diff estimates at the 1st choice margin for pulled-up and pushed-down groups. The RD estimates are more similar to the results for pushed-down students (but smaller for graduation from the initial admission), with the same sign and order of magnitude.

In addition, I estimate the same RD model while limiting the sample to students with a boost score greater than 5 (the average boost score for pushed-down students at the margin of first and no admission) in an effort to recover the effects from a population that is more comparable to the pulled-up group of students. Table B.27 shows the enrollment results, while Table B.27 displays the graduation estimates. In both instances, the outcomes are greater and comparable to the diff-in-diff outcomes (19.5 p.p. for enrollment and 9.9 p.p for graduation).

The tables B.33 and B.33 provide the findings of an alternative exercise designed to quantify the effects on a subset of pulled-up students. This experiment focuses on students with admission at their 1st or 2nd preference with the relative GPA (treatment margins 1-2) and with simulated admission at their 2nd preference without the GPA$^+$ measure. In this sample, the threshold crossing is only explained by the boost). By comparing pulled-up students with very comparable unaffected students, the sample restriction aims to determine the effect of threshold crossing for pulled-up students (non-crossing but similar - close to the margin). The small sample size resulting from the requisite makes the results unstable and imprecise; still, the sign and magnitude of the values for graduation from the admission offer fluctuate around the diff-in-diff estimate for the margin between first and second preference (7.7 p.p).

A potential threat to the regression discontinuity (RD) design is that people might try to sort themselves above the cutoff in order to receive an offer from their preferred program. Figures in Appendix B.1 show that there are no discontinuities around the cutoffs in the density of applicants and in the observed characteristics support the assumption against that type of sorting. In addition, the McCrary (2008) test is negligible and fails to reject the null hypothesis of no sorting.

2.9 Robustness checks

I conduct a number of checks to verify the robustness of my conclusions. I check different samples (removing students with boost higher than 150 points or students attending programs over 6 years) and estimating my results clustering at the school-year level, and all of them support my main findings.

Changes in ROL due to the reform The key assumption for the identification of pulled-up and pushed-down groups is that the rank order list (ROL) of the application submitted by the applicants in each process would not change under a different assignment mechanism. Recent literature presents evidence raising concerns over the inclusion of more selective programs when the boost score is observed. By checking the selectivity of the first preference listed by students in 2012 and 2013 (measured as the cutoff score of that program) for students with the same boost we see some increase in the selectivity when the boost is larger than 150.

As a robustness check, I estimate the main results presented above but remove students with boost scores higher than 150. The tables with the results for this case are presented in Appendix B.1. Results are not only qualitative but also quantitative similar for all the outcomes.

Sensitivity of the results to long programs Given the instability of graduation results even after 8 years of participation in the admission process, I restrict the analysis only to programs with an expected graduation time of less than 6 years in Appendix B.1 and to less than 7 years in Appendix B.1. Both sets of results present similar results in terms of magnitude and significance than the ones discussed previously.

Inference The previous results have been estimated using robust standard errors. Alternately, in Appendix B.1 I present the main results allowing clustering at the school-year level. Nonetheless, any of the results take into consideration the potential error associated with the estimation of the pulled-up and pushed-down groups. Results presented in Appendix B.1 are virtually equivalent to the results presented above.

2.10 Labor market outcomes

Finally, I study the labor market effects of the reform.[36] An important challenge refers to the long graduation times observed in the previous section, and the even longer span of time needed to account for the behavioral responses of reapplication to the selective system when students were not satisfied by their admission offer. Therefore, by studying earnings ten years after the implementation of the reform I am not able to fully capture the effect of the reform on earnings, limiting the analysis. Moreover, aggregated data - earnings with an indicator of a group of treatment but without individual characteristics- only allows for very preliminary evidence at the group level.

Figure 2.5 presents earnings histograms for pulled-up and pushed-down groups of students pre and post-implementation of the relative GPA reform. In each case, histograms are presented relative to the unaffected group. Even though at the moment I cannot calculate the diff-in-diff estimates, a preliminary review of the aggregated data confirms that pulled-up and pushed-down students do not do worse than before

[36]Up to this date, access to individual-level data required to estimate the difference-in-differences specification used in the previous sections is under approval.

the implementation of the reform. Overall, in terms of outcome efficiency - graduation and earning-, the evidence confirms that the new assignment mechanism didn't make the system less efficient.

2.11 Conclusion

This paper studies the impact of providing students with access to more selective college alternatives. I use the variation on admission generated by the inclusion of a relative GPA measure motivated by equity concerns. I explore the effects of the reform on enrollment, graduation, and earnings for the two groups directly and indirectly affected by this change: (i) students who gain access to more selective programs (pulled-up) and (ii) students who lose access to more selective programs (pushed-down).

The transparency of the college admission process combined with the properties of the assignment mechanism and the richness of the data available allows me to cleanly identify the groups of students affected by the reform, one of the big challenges in the evaluation of admission reforms. By simulation of the admission offers with and without the inclusion of the relative GPA measure I identify the group of affected students. The replication of the admissions with the GPA^{+} in the years before the reform helps me to identify the group of students that would have been affected. This simulation facilitates the implementation of a difference-in-difference design.

This empirical strategy compares the outcomes of students in the pulled-up and pushed-down groups before and after the implementation of the reform, therefore, before and after they get access to these more selective programs. The transitory variation in outcomes is controlled by the second difference with respect to the group of unaffected students.

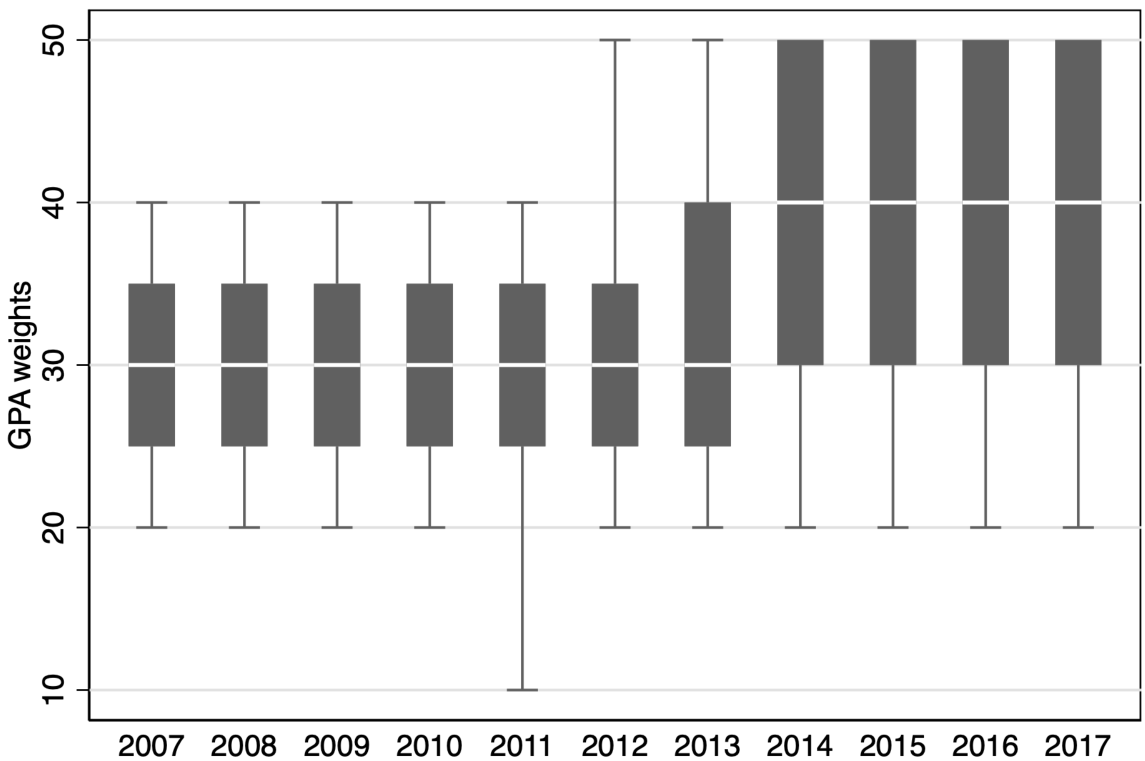
I find that the incorporation of the relative GPA measure into the college admissions application score formula expanded the options available for students with significantly less resources. As a result of the reform, pulled-up students became more likely to enroll in a selective program, and they chose to enroll in programs where their peers have higher test scores, GPA scores, and graduation rates. Contrary to the prediction of the mismatch hypothesis, reform-targeted applicants with lower test scores gained from enrolling in more selective options, boosting their likelihood of graduation by 8.4 percentage points.

For pushed-down students, I find that their likelihood of graduating from the admission program assigned by the new mechanism decreases by 8.2 p.p., but they are not less likely to receive a bachelor's degree. There is however an impact in the timing of their enrollment that would be interesting to study with more details once more data on graduation and earning becomes available. Nevertheless, preliminary evidence confirms that there is no negative impact on earnings for pushed-down students.

Collectively, the evidence presented above indicates that test-based meritocratic admission system can be improved by the inclusion of in-school performance metrics, increasing admission equity without incurring an efficiency penalty.

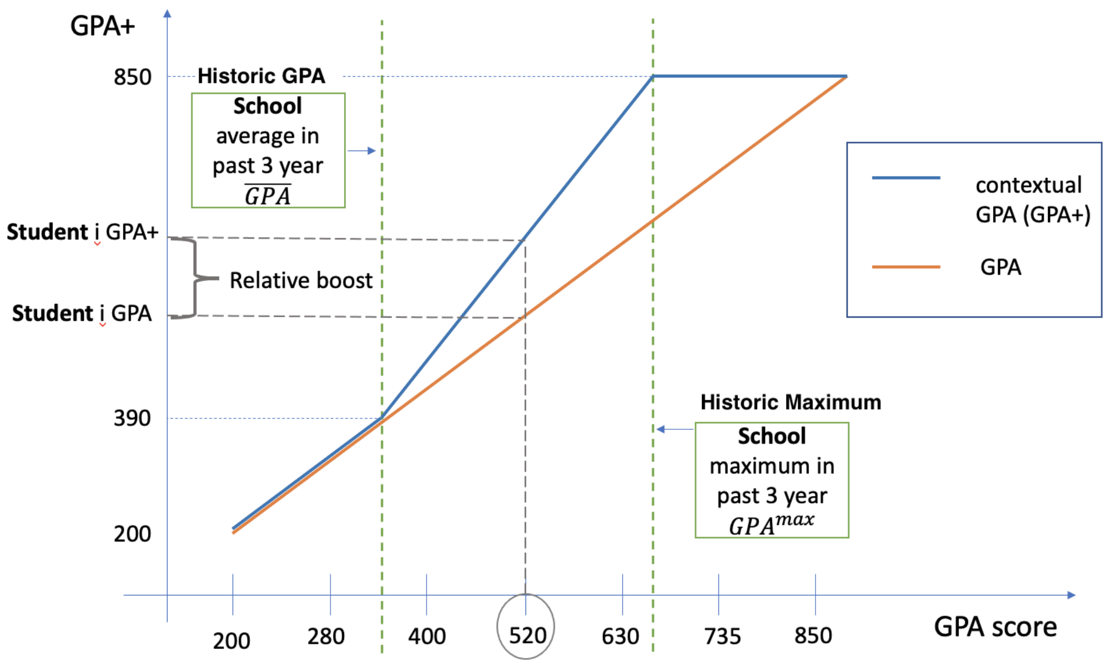
2.12 Figures and tables

Figure 2.1: Change in weights of GPA components (GPA + GPA^{+}) in application scores formula by year



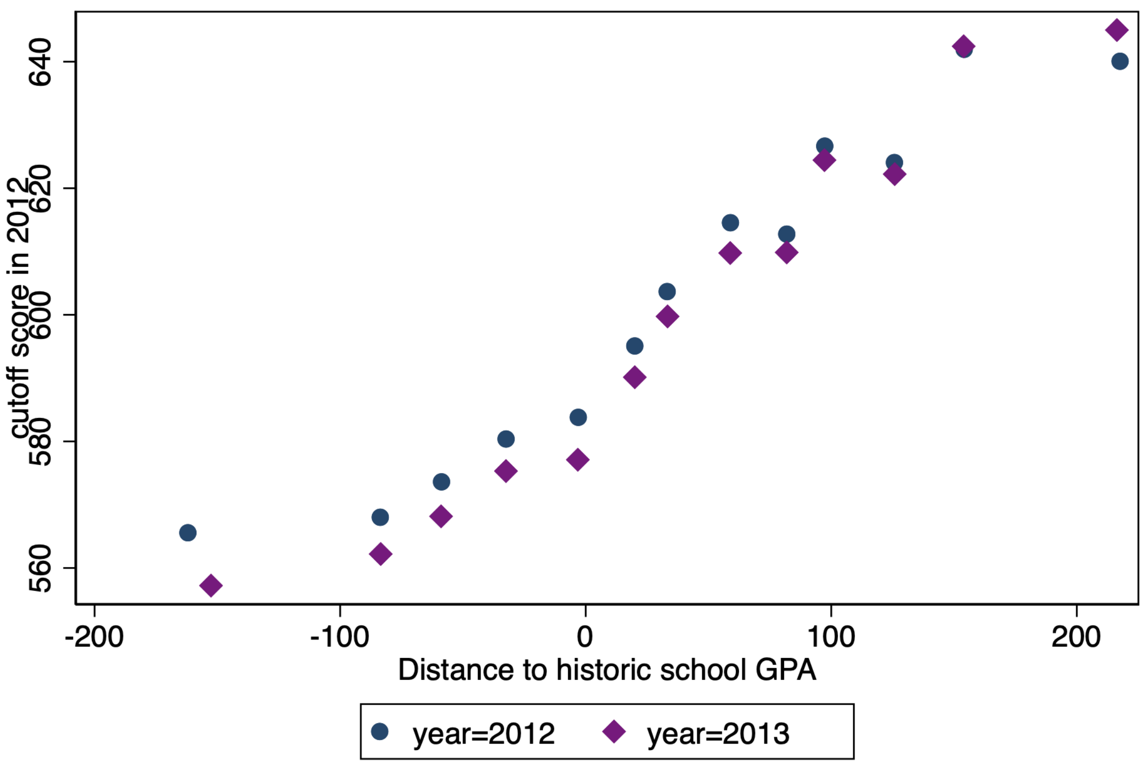
Notes: This figure shows the whisker plots for the distribution of the weights of the GPA components assigned by programs in the application score formula. The middle box represents 50% of the data, the white line corresponds to the median weight, and the maximum and minimum values are displayed with vertical lines ("whiskers").

Figure 2.2: Example boost score



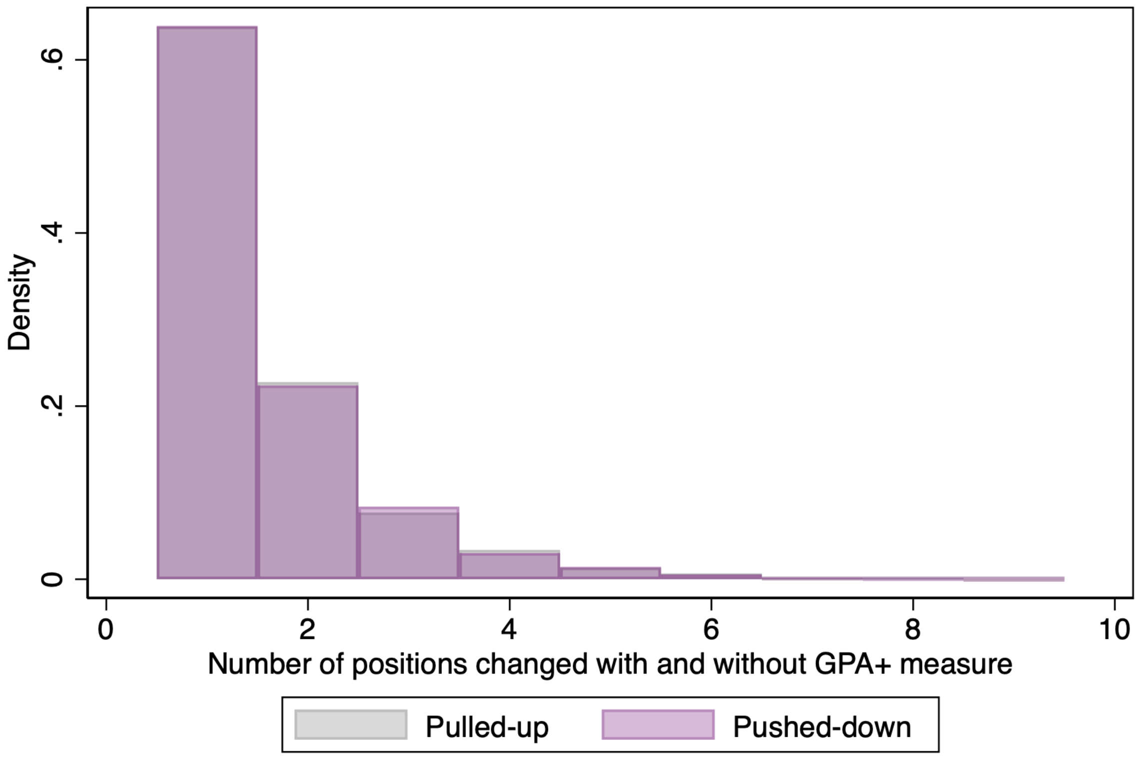
Notes: exemplary figure to show how GPA^+ depends on school averages and how it relates to the GPA score. Boost is obtained from the difference between GPA^+ score and GPA.

Figure 2.3: Selectivity of program ranked 1st by relative position of student at their high school



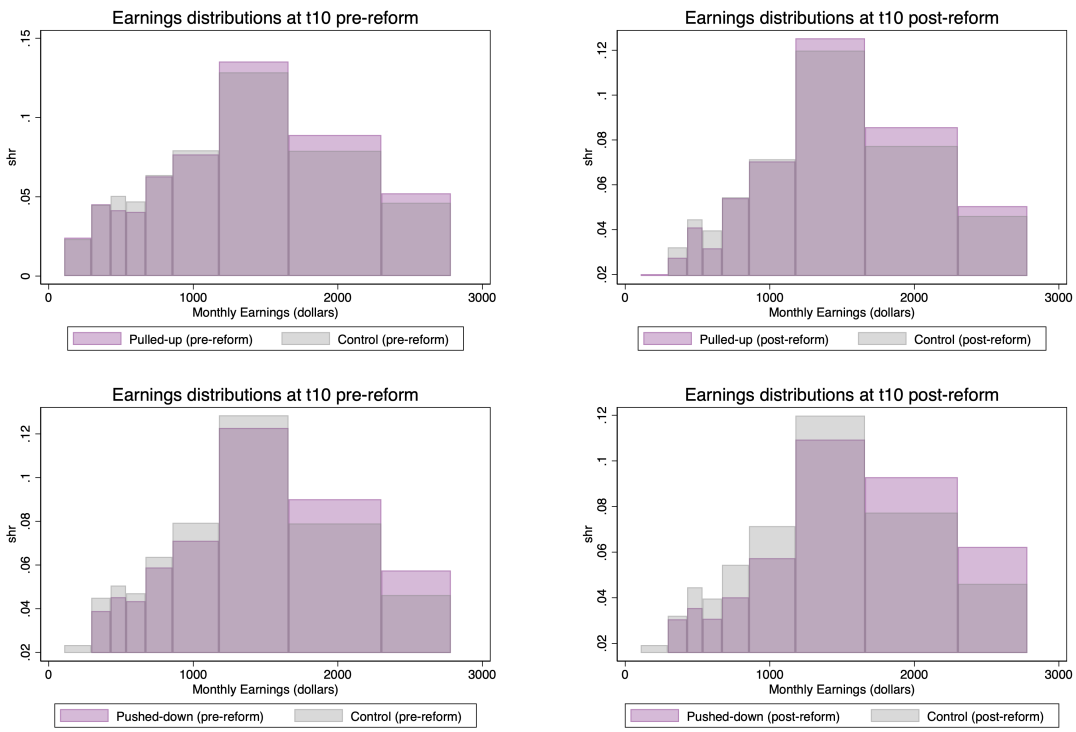
Notes: binscatter of the selectivity of the 1st preference by boost. Selectivity is measured as the cutoff (application score of the last person admitted in the programs, measured pre-reform) of the program listed 1st. The x-axis has the GPA^+ measure but is centered around the average score of the school. By centering on the school average we have that positive values correspond to the boost score.

Figure 2.4: Distribution of the change of the preference of admission in the ranking of preferences



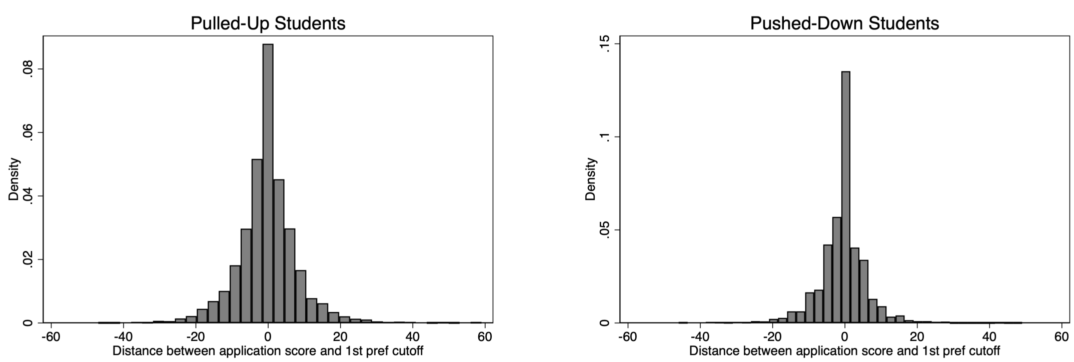
Notes: distribution of pulled-up and pushed-down students based on the number of positions moved in their ranking between admission with and without GPA^{+}.

Figure 2.5: Earnings distribution



Notes: Earnings distribution for pulled-up and pushed-down groups, relative to unaffected, 10 years after application. Figures on the left show earnings distribution for students in cohort 2012 (pre-reform) and figures on the right show earnings distribution for students in cohort 2013 (post-reform).

Figure 2.6: Distribution by distance between application score and cutoff for 1st preference



Notes: Histogram for pulled-up and pushed-down groups for students who win or lose their first preference when the relative GPA is considered.

Table 2.1: Distribution of students reporting rankings by year

Ranking		2012	2013
1	Total N	116,336	118,208
	Only 1 (%)	0.07	0.07
	Up to 1 (%)	0.07	0.07
2	Total N	108,715	110,264
	Only 2 (%)	0.09	0.10
	Up to 2 (%)	0.16	0.17
3	Total N	98,166	98,245
	Only 3 (%)	0.17	0.20
	Up to 3 (%)	0.32	0.37
4	Total N	78,828	74,152
	Only 4 (%)	0.16	0.17
	Up to 4 (%)	0.48	0.55
5	Total N	60,420	53,693
	Only 5 (%)	0.14	0.14
	Up to 5 (%)	0.62	0.68
6	Total N	44,322	37,403
	Only 6 (%)	0.10	0.09
	Up to 6 (%)	0.72	0.78
7	Total N	32,720	26,182
	Only 7 (%)	0.07	0.07
	Up to 7 (%)	0.79	0.84
8	Total N	24,208	18,477
	Only 8 (%)	0.06	0.05
	Up to 8 (%)	0.85	0.89
9	Total N	17,041	12,572
	Only 9 (%)	0.04	0.03
	Up to 1 (%)	0.89	0.92
10	Total N	12,582	9,167
	Only 10 (%)	0.11	0.08
	Up to 10 (%)	1.00	1.00

Notes: the table shows the total number of students reporting each ranking, the percentage of students reporting a total of each ranking, and the percentage of students reporting each ranking or fewer options. In 2011 the maximum number of choices was increased and students, in the 2012 process, were nudged to take advantage of that and list 10 options.

Table 2.2: Summary Statistics for Groups of Interest

	Unaffected		Pulled-up		Pushed-down	
	2012	2013	2012	2013	2012	2013
N	108,167	109,440	3,753	4,515	4,416	4,253
Female (%)	53	52	62	60	41	40
Public School (%)	28	27	29	29	26	25
Voucher School (%)	53	54	60	60	47	47
Private School (%)	19	18	10	11	27	28
Family Inc ($/mo)	689	714	573	594	809	869
Father with HS (%)	67	67	64	61	74	75
Mother with HS (%)	73	73	69	70	78	79
Father with College (%)	26	26	20	19	34	35
Mother with College (%)	21	21	16	16	27	29
Capital City (%)	39	39	46	46	54	53
Std Math	0.68	0.65	0.74	0.65	1.05	1.13
Std Verbal	0.66	0.65	0.70	0.62	1.01	1.04
Std GPA	0.75	0.73	1.40	1.28	0.42	0.58
Boost score	21	22	60	57	6	8

Notes: This table shows the summary statistics for the groups of interest, in the year before and after the reform.

Table 2.3: Distribution of pulled-up students by fields with and without GPA$^+$

With GPA$^+$	Without GPA$^+$									
Field	MedOdon	Health	Sci	Engi	Tech	Business	Art	SocSci	Law	Educ
MedOdon	126	54	7	10	2	2	0	3	4	1
Health	3	425	51	21	34	21	2	32	2	44
Sci	0	19	57	23	25	3	0	5	0	10
Engi	1	7	23	359	111	51	2	5	2	2
Tech	0	9	19	79	266	20	20	7	1	10
Business	0	1	12	19	29	224	3	8	3	6
Art	0	0	1	0	7	0	18	7	0	4
SocSci	0	8	8	3	17	32	15	200	18	47
Law	0	0	0	1	4	11	0	29	63	9
Educ	0	4	4	6	13	9	6	29	4	194

Notes: Total number of pulled-up students in 2013 in each field combination based on the field of the program that they get admitted with the GPA$^+$ and the field of the program that they get admitted without the GPA$^+$.

Table 2.4: Distribution of pushed-down students by fields with and without GPA^{+}

With GPA^{+} Ranking	Without GPA^{+} MedOdon	Health	Sci	Engi	Tech	Business	Art	SocSci	Law	Educ
MedOdon	126	54	7	10	2	2	0	3	4	1
Health	3	425	51	21	34	21	2	32	2	44
Sci	0	19	57	23	25	3	0	5	0	10
Engi	1	7	23	359	111	51	2	5	2	2
Tech	0	9	19	79	266	20	20	7	1	10
Business	0	1	12	19	29	224	3	8	3	6
Art	0	0	1	0	7	0	18	7	0	4
SocSci	0	8	8	3	17	32	15	200	18	47
Law	0	0	0	1	4	11	0	29	63	9
Educ	0	4	4	6	13	9	6	29	4	194

Notes: Total number of pushed-down students in each field combination based on the field of the program that they get admitted with the GPA^{+} and the field of the program that they get admitted without the GPA^{+}.

Table 2.5: Difference-in-differences estimates for 2012 and 2011

	(1) Enroll	(2) Enroll	(3) Grad by 8yr	(4) Grad by 8yr
Pulled-Up	0.001	-0.015	0.006	-0.009
	(0.011)	(0.010)	(0.011)	(0.011)
Pushed-Down	-0.010	-0.002	0.007	0.005
	(0.009)	(0.009)	(0.011)	(0.010)
Observations	211,872	211,872	211,872	211,872
Controls		✓		✓

Notes: columns 1 and 3 have the estimates from the difference-in-difference without controls and columns 2 and 4 have the estimates for the same outcomes but controlling by individual characteristics. Robust standard errors are in parentheses.

Table 2.6: Diff-in-diff estimates for enrollment

	(1) Enrollment	(2) Enrollment	(3) Non-Select	(4) Non-Select	(5) Enrollment	(6) Enrollment
Pulled-Up x after	0.199***	0.219***	-0.049***	-0.057***	0.165***	0.175***
	(0.011)	(0.010)	(0.007)	(0.007)	(0.0113)	(0.0111)
Pushed-Down x after	-0.136***	-0.167***	0.023***	0.039***	-0.0947***	-0.110***
	(0.010)	(0.010)	(0.007)	(0.006)	(0.00987)	(0.00965)
Obs.	234,544	234,544	234,544	234,544	186,734	186,734
Controls		✓		✓		✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.000	0.000	0.008	0.048	0.000	0.000

Notes: columns 1 and 2 have estimates when the outcome is enrollment at the admission program. Columns 3 and 4 have estimates for an indicator if the student enrolls in a non-selective program. Columns 2 and 4 control for standardized test scores, GPA, family income, region, type of high school, and gender. Columns 5 and 6 restrict the sample to students with some admission offers under both regimes to capture the enrollment effect on students at the intensive margin. Robust standard errors are in parentheses.

Table 2.7: Changes in peer characteristics at chosen programs

Program	Diff-in-Diff		Pre-Reform ($\bar{x}$)		
Charact.	Pulled-up	Pushed-down	Control	Pulled-up	Pushed-down
Math (std)	0.264***	-0.212***	1.104	1.262	1.272
	(0.008)	(0.007)	[0.610]	[0.603]	[0.611]
Verbal (std)	0.235***	-0.230***	1.114	1.240	1.261
	(0.008)	(0.008)	[0.567]	[0.537]	[0.530]
GPA (std)	0.280***	-0.288***	1.165	1.317	1.276
	(0.009)	(0.008)	[0.575]	[0.519]	[0.544]
Grad on time	0.044***	-0.040***	0.389	0.384	0.374
	(0.007)	(0.006)	[0.263]	[0.271]	[0.267]
E(grad time)	0.026	-0.065**	5.110	5.163	5.161
	(0.027)	(0.027)	[0.725]	[0.779]	[0.820]

Notes: Columns 1 and 2 show the results for the main diff-in-diff specification for the outcome 5 different outcomes: (i) average math score of students enrolled at the chosen program pre-reform, (ii) average verbal score of students enrolled at the chosen program pre-reform, (iii) average GPA score of the students enrolled at the chosen program pre-reform, (iv)probability of graduation on time by the students enrolled at the chosen program pre-reform, (v) expected graduation time based on the class structure at the chosen program. Columns 3-5 show the averages and standard deviation of these variables for the 3 groups of interest, pre-reform. Robust standard errors are in parentheses.

Table 2.8: Effect on re-application by second year

	(1) Reapplication	(2) Reapplication
P-Up x after	-0.0387***	-0.0374***
	(0.0091)	(0.0091)
P-Down x after	0.0739***	0.0717***
	(0.0086)	(0.0086)
Obs.	234,544	234,544
Controls		✓

Notes: diff-in-diff estimates using an indicator if the student participates in the application process in the second year. Robust standard errors are in parentheses.

Table 2.9: Effect on total enrollment up to 4 years after the reform

	(1) Enroll at t=1	(2) Enroll at t=2	(3) Enroll at t=3	(4) Enroll at t=4
P-Up x after	0.062***	0.018**	0.015*	0.003
	(0.008)	(0.008)	(0.008)	(0.009)
P-Down x after	-0.077***	-0.008	0.005	0.008
	(0.007)	(0.007)	(0.008)	(0.008)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓

Notes: diff-in-diff estimates for an indicator if the student is enrolled in some program at different points in time. Column 1 shows the results for the same year of applications; column 2 for two years after the application process; column 3 for three years after and column 4 for four years after. Robust standard errors are in parentheses.

Table 2.10: Effect on peers test scores at enrollment up to 4 years after the reform

	(1) Selectivity at t=1	(2) Selectivity at t=2	(3) Selectivity at t=3	(4) Selectivity at t=4
P-Up x after	0.160***	0.093***	0.117***	0.107***
	(0.008)	(0.009)	(0.010)	(0.012)
P-Down x after	-0.120***	-0.113***	-0.084***	-0.082***
	(0.008)	(0.008)	(0.010)	(0.011)
Obs.	187,534	190,703	181,081	175,037
Controls	✓	✓	✓	✓

Notes: diff-in-diff estimates for average test score (standardized) of students at program chosen by applicants at different points in time. Columns 1 through 4 show the results from 1st to 4th year since the moment of application. Robust standard errors are in parentheses.

Table 2.11: Effect on graduation from initial admission program

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.042***	0.072***	0.084***	0.043***
	(0.008)	(0.009)	(0.010)	(0.010)
P-Down x after	-0.033***	-0.060***	-0.082***	-0.039***
	(0.007)	(0.009)	(0.009)	(0.010)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.397	0.366	0.917	0.742

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table 2.12: Effect on graduation from initial admission program conditional on some admission offer with both mechanism

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.037***	0.065***	0.078***	0.034***
	(0.009)	(0.011)	(0.011)	(0.012)
P-Down x after	-0.020**	-0.040***	-0.062***	-0.017
	(0.009)	(0.010)	(0.011)	(0.011)
Obs.	186,734	186,734	186,734	186,734
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.195	0.100	0.305	0.295

Robust standard errors in parentheses

Notes: Diff-in-diff results for the sample of students with some admission with and without the inclusion of the GPA$^+$ measure. Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admitted program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table 2.13: Mismatch effect exercise

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.047***	0.078***	0.091***	0.049***
	(0.007)	(0.009)	(0.010)	(0.010)
Obs.	234,529	234,529	234,529	234,529
Controls	✓	✓	✓	✓

Robust standard errors in parentheses

Notes: This table shows the effect on graduation from admission into a more selective program. The diff-in-diff specification controls by the admission program without the relative GPA reform in order to ensure that estimation uses only variation from students with admission to more selective programs after the reform, and not from potential changes in the compositions of admission programs between 2012 and 2013. Robust standard errors are in parentheses.

Table 2.14: Effect on STEM applicants

	(1) Admission	(2) Enrollment	(3) Enrollment	(4) Grad by 8yr	(5) Grad or enroll by 8 yr
P-Up x after	0.061***	0.216***	0.169***	0.061***	0.052***
	(0.010)	(0.014)	(0.015)	(0.014)	(0.016)
P-Down x after	-0.030***	-0.166***	-0.120***	-0.047***	-0.035**
	(0.010)	(0.013)	(0.012)	(0.014)	(0.016)
Obs.	234,544	110,791	97,350	97,350	97,350
Controls	✓	✓	✓	✓	✓

Notes: Column 1 shows the coefficient for the indicator of admission offer in STEM using the main diff-in-diff specification. Column 2 shows the effect on enrollment for STEM applicants. Column 3 restricts the sample of column 2 only to students that have some admission offer with and without GPA. Column 4 presents the effects on graduation for the same sample as column 3. Finally, column 5 presents the results for graduation or still enroll in a STEM program. Robust standard errors are in parentheses.

Table 2.15: Effects on graduation from any program

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8 yr
P-Up x after	-0.003	0.005	-0.008	0.006
	(0.009)	(0.011)	(0.011)	(0.010)
P-Down x after	-0.012	-0.015	-0.032***	-0.006
	(0.009)	(0.010)	(0.011)	(0.010)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓

Notes: diff-in-diff estimates for the indicator if the student graduate from some program by 6, 7, or 8 years. Column 4 show the results when the dependent variable takes the value of 1 if the student graduate or if the student is enrolled in some program 8 years after application. Robust standard errors are in parentheses.

Table 2.16: Effect on graduation from a selective program

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8yr
P-Up x after	0.019**	0.030***	0.019*	0.024**
	(0.009)	(0.010)	(0.011)	(0.010)
P-Down x after	-0.022***	-0.028***	-0.046***	-0.030***
	(0.008)	(0.010)	(0.010)	(0.010)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓
	Selective	Selective	Selective	Selective

Robust standard errors in parentheses

Notes: Columns 1-3 show the results for graduation from a selective program by 6, 7, or 8 years after application. Columns 4-5 show the same results for non-selective programs. Robust standard errors are in parentheses.

Table 2.17: Effects on graduation from a non-selective program

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8yr
P-Up x after	-0.022***	-0.025***	-0.026***	-0.019***
	(0.004)	(0.005)	(0.005)	(0.006)
P-Down x after	0.009***	0.013***	0.014***	0.022***
	(0.004)	(0.004)	(0.004)	(0.005)
Obs.	234,544	234,544	234,544	234,544
Controls	✓		✓	✓
	Non-Selective	Non-Selective	Non-Selective	Non-Selective

Notes: Columns 1-3 show the results for graduation from a selective program by 6, 7, or 8 years after application. Columns 4-5 show the same results for non-selective programs. Robust standard errors are in parentheses.

Table 2.18: Differential effect for students with big and small changes in selectivity

	(1) Enrollment	(2) Grad by 8yr	(3) Grad by 8yr from any	(4) Grad or enroll by 8 yr
Small Pulled-Up x after	0.195***	0.069***	-0.034**	0.005
	(0.015)	(0.016)	(0.017)	(0.015)
Big Pulled-Up x after	0.151***	0.080***	0.007	-0.003
	(0.016)	(0.016)	(0.017)	(0.015)
Small Pushed-Down x after	-0.108***	-0.040**	0.000	0.007
	(0.013)	(0.016)	(0.017)	(0.015)
Big Pushed-Down x after	-0.112***	-0.085***	-0.036**	-0.014
	(0.014)	(0.016)	(0.017)	(0.015)
Obs.	186,734	186,734	186,734	186,734
Controls	✓	✓	✓	✓

Notes: Based on how much the average test score of the peers (selectivity of the programs) changed between the simulated program and the admission program, the pulled-up and pushed-down groups are split into big and small changes in selectivity. The sample contains only students who have some admission offer in both regimes. Column 1 shows how the reform affected enrollment for the four subgroups. Column 2 shows the change in the probability of program completion. Column 3 shows the effects of graduating from any program 8 years after admission. Column 4 shows the effects of graduating or still being in school 8 years after application, which takes into account the fact that students in the selective system may switch programs, which will cause them to graduate from college later. Robust standard errors are in parentheses.

Table 2.19: Effects on students move into or out of their 1st preference with and without GPA^+

	(1) Enroll	(2) Select	(3) Grad by 8yr	(4) Grad (any) by 8yr	(5) Grad or enroll by 8 yr
P-Up x after	0.281***	0.283***	0.112***	-0.005	-0.002
	(0.014)	(0.007)	(0.014)	(0.015)	(0.014)
P-Down x after	-0.179***	-0.228***	-0.101***	-0.026*	0.003
	(0.013)	(0.007)	(0.015)	(0.016)	(0.014)
Obs.	225,618	225,531	225,618	225,618	225,618
Controls	✓	✓	✓	✓	✓

Robust standard errors in parentheses

Notes: Column 1 presents the results for enrollment at their admission offer. Column 2 presents the estimates for the change in the average test scores at the program of enrollment. Columns 3 to 5 show the estimates for graduation from the admission offer, from any program, and for an indicator of graduating or still enrolled 8 years after the application process. Robust standard errors are in parentheses.

Table 2.20: Difference-in-differences and RD comparison

	(1) RD	(2) DD Pulled-Up	(3) DD Pushed-Down
Enrollment	0.178***	0.281***	-0.179***
	(0.006)	(0.014)	(0.013)
Selectivity	0.16***	0.157***	-0.168***
	(0.005)	(0.010)	(0.010)
Grad from Admission	0.067***	0.112***	-0.101***
	(0.007)	(0.014)	(.015)
Grad from Any	-0.004	-0.005	-0.026*
	(0.007)	(0.015)	(0.016)
Graduation or Enroll	-0.004	-0.002	0.003
	(0.007)	(0.014)	(0.014)
Observations	118,205	225,618	225,618
Controls	✓	✓	✓

Notes: Column 1 presents the results for the RD specification. It compares people with very similar application scores for their 1st preference in 2013. Column 2 presents the results for pulled-up group with the diff-in-diff specification. It compares students who are above the threshold for their 1st choice to their "past cohort selves" who did not get their 1st choice. Their "past cohort selves" have the same score, presumably not too far below the threshold, but did not get treated. Column 3 presents the results for pushed-down group with the diff-in-diff specification. It compares people who are below the 2013 threshold to their "past cohort selves" who did get their 1st choice. Robust standard errors are in parentheses.

Chapter 3

Gender bias in college admissions based on test scores: evidence and policy recommendations

3.1 Introduction

Gender sorting into majors and occupations has shown to be important for explaining labor market disparities between men and women (Blau and Kahn; Huneeus et al.; Sloane, Hurst, and Black). The causes of these sorting patterns, however, remain contentious in the literature. One narrative relies on preferences (e.g., Croson and Gneezy), however, Bertrand argues they cannot rationalize observed wage gaps in the labor market. Goldin flags the importance of industry-specific returns to certain types of working schedules while recent papers document additional gender disparities in the sorting process, for example, gender-specific returns to majors (Zimmerman; Aguirre, Matta, and Montoya) or gender gaps in peer recognition, referrals, and promotions (Sarsons; Benson, Li, and Shue; Card et al.; Cullen and Perez-Truglia; Haegele; Exley and Kessler). In this context, the understanding of the *sources* of gender sorting and its policy implications remains an area of active research.

In this paper, we explore the role of standardized test scores in the context of college admissions on gender sorting into majors and occupations. The (sometimes contested) rationale for using standardized test scores in college admission processes is rooted in the idea that standardized tests allow fair comparisons between students based on their ability, which in turn predict some notion of *latent performance* in college. Even though standardized tests are usually designed to be objective and unbiased, the predictive ability of the test may vary by gender. Then, if standardized tests systematically under-predict the ability of women relative to men, then female applicants will face stricter *de facto* selection thresholds compared to their male counterparts because, at the margin, they will need higher latent performance indexes to meet the program selection thresholds. As a consequence, relying on standardized tests may imply that women are less likely to be accepted either in college or in their most preferred programs than men with similar latent performance, generating misallocation in the college application process. This misallocation may affect women's presence in selective programs and, potentially, their long-run labor market outcomes.

Testing for gender bias in the predictive ability of standardized tests, however, has proven to be challenging for at least two reasons. First, many college selection processes are decentralized and do not have explicit rules for test scores. For example, colleges may have autonomy in determining their admissions policies in a way that is unobserved to researchers and may complement (or even substitute) test scores with alternative selection instruments. Then, it is difficult to isolate the effect of standardized tests in the allocation process. Second, it is difficult to gather data that links the applications with both enrollment and ex-post measures of performance, which could be used as proxies of the unobserved latent variable that the college admission process is seeking to predict.

We overcome these challenges and explore for gender differences in the predictive ability of standardized tests by using nationwide administrative records from Chile. We leverage detailed student-level data on the application process (including rank-ordered preferences and enrollment) and college performance (first-year GPA and graduation rates), as well as other demographic characteristics of the applicants. As we describe in Section 3.3, Chile is an appealing setting to test the proposed mechanism since it has a centralized college admission system that uses a textbook deferred-acceptance algorithm that is exclusively based on applicants' academic performance – measured by test scores from a national standardized entrance exam combined with high school outcomes – and their preferences for programs – which are a combination of college and major.

The analysis proceeds in two steps. The first step builds on the literature of marginal outcome tests developed in the discrimination literature (Hull). This model considers an ideal assignment rule to programs based on an unobserved latent outcome, Y^*, which in our setting is proxied by first-year GPA and graduation rates. Since the outcome is ex-ante unobserved, the assignment rule is based on its prediction, $\hat{Y}$, which in our setting coincides with the standardized test score. If the prediction error varies between demographic groups, then the group with worse prediction errors is less likely to make the threshold to get admitted to a program relative to students with the same latent outcome, which can be interpreted as a source of bias in admissions. In other words, test scores are considered biased against women if the systematic differences in prediction errors lead to systematic differences in effective thresholds and, therefore, to systematic disparities in college admissions. Since we directly observe the applicants' scores, the nominal thresholds, and proxies for the ex-post latent outcome, we can exactly compute the differences at the margin, where the bias has the potential to concretely impact the application result. This differs from other applications where the selection decision is idiosyncratically taken by (possibly heterogeneous) decision-makers and, therefore, marginal individuals have to be identified using additional methods (e.g., Arnold, Dobbie, and Yang; Arnold, Dobbie, and Hull; Grau and Vergara).

We find that standardized tests systematically under-predict college performance for marginally selected female applicants relative to their male counterparts. We estimate significant gender differences in outcomes between females and males at the margin of admission: 1st-year of college GPA is larger for marginally accepted female applicants, and marginally accepted female applicants are 7 percentage points more likely to graduate from the program they enter. Our results are consistent with women facing selection thresholds that are, on average, three points larger estimated from the 1st year college GPA and from the probability of graduation. Then, through the lens of a model of bias in the selection process, these results suggest that, despite being designed to be gender-neutral, standardized tests are biased against female applicants, echoing the notion of institutional discrimination developed in Bohren, Hull, and Imas.

With the bias estimations in hand, the second step performs three counterfactual exercises to assess the consequences, in terms of admissions, of the gender differences in the performance predictability of standardized test scores. The first exercise – "partial equilibrium" – computes how many male applicants would not have been admitted to the program they enrolled in if faced with the effective threshold of female applicants (three points larger than the official cutoff). The second and third exercises – "general equilibrium" – simulate admissions following the deferred acceptance mechanism used in the Chilean admission system by perturbing the inputs of the algorithm. The second exercise gives three extra points to female applicants, mimicking a case in which the system compensates for the bias in the standardized test score. In the third exercise, we gradually increase the weights assigned to the high school GPA components in the formula to calculate the application score, which we find to be negatively correlated with the size of the bias.

The first exercise suggests that around 50% of programs would lose between 1 and 7 male students if they were faced with the female effective threshold. The second exercise suggests that around 2% of female applicants would be better off with the proposed changes by accessing a more preferred option. Since these exercises hold open seats fixed, the improvement for female applicants comes at the cost of male applicants. The third exercise suggests that for assignments of weights to the high school GPA component of 40% or 50% it is possible to improve the assignment outcomes of around 2% of female applicants. However, this result differs from the second exercise in part due to a higher number of students with their admission changed, i.e.more males and more females with more preferred and less preferred admission than in the alternative intervention. Taking it all together, this implies that the consequences of the differential predictive ability in terms of misallocation are not negligible.

Related literature

This paper contributes to different literature. First, it contributes to the understanding of gender gaps in education and labor markets. For surveys on gender gaps in education, see Buchmann, DiPrete, and McDaniel, Bertocchi and Bozzano, and Evans, Akmal, Jakiela, et al. For surveys on gender gaps in the labor market, see Altonji and Blank, Bertrand, and Blau and Kahn. Scholars have documented systematic gender gaps in test scores with men having better performances in math but worst performances in reading (Niederle and Vesterlund; Pope and Sydnor; Hermann and Kopasz). Our analysis differs from this literature since it focuses on gender gaps in performance conditional on test scores, allowing us to quantify biases in college admissions. While gender gaps in educational attainment have decreased over time, they remain important in tertiary education, especially in developing countries and certain fields. Our analysis suggests that the differential predictive ability of standardized tests plays a role in explaining these gaps. It also suggests that exclusively looking at attainment gaps may be misleading since there are also gaps in the allocation to programs conditional on attending college. Given the returns to college selectivity and major choice (Card; Dale and Krueger; Dale and Krueger; Hout; Zimmerman; Kirkeboen, Leuven, and Mogstad; Bleemer and Mehta; Chetty et al.; Black, Denning, and Rothstein; Mountjoy), our analysis provides an additional rationale for gender gaps in labor markets.

This paper also contributes to the analysis of college admission systems. The optimal policy mix between standardized tests and alternative policies such as affirmative action or top percent policies remains an active area of discussion given the lack of consensus about the extent to which different policies meet different equity and efficiency objectives (Card and Krueger; Fryer Jr and Loury; Krueger, Rothstein, and Turner; Rothstein and Yoon; Black, Cortes, and Lincove; Card and Giuliano; Yagan; Hyman; Kapor et al.; Bleemer; Ellison and Pathak; Mello; Otero, Barahona, and Dobbin; Rothstein). In this context, our paper contributes to the documentation of problems in the predictive accuracy of standardized tests (Rothstein; Black, Cortes, and Lincove; Jacob and Rothstein; Mattern, Sanchez, and Ndum; Mattern et al.; Marini et al.; Westrick et al.), in particular, on gender differences and their impact on admission outcomes.

Finally, our results add to the literature of bias in presumably objective selection instruments which mainly focuses on algorithmic decision-making in high-stakes environments such as healthcare and the criminal justice system (Angwin et al.; Hardt, Price, and Srebro; Kleinberg, Mullainathan, and Raghavan; Kleinberg et al.; Arnold, Dobbie, and Hull; Berk et al.; Mullainathan and Obermeyer; Rambachan; Rambachan et al.). Our results suggest this literature could be extended to the design of the optimal policy mix in college admission systems to attenuate the concerns on existing instruments.

3.2 Theoretical framework

This section describes the theoretical framework that guides the empirical analysis. It makes explicit the notion of bias we use in this paper and discusses the assumptions needed for its identification using marginal outcome tests, which we empirically implement in Section 3.4.

Model Let i index applicants and j programs, which define combinations of colleges and majors. The system seeks to assign applicants to programs based on some latent outcome, Y^*_{ij}. Below we discuss the interpretation of Y^*_{ij} with more detail. Y^*_{ij} is unobserved at the time of the application but ex-post measurable for selected individuals. Therefore, the application process is based on predictions of Y^*_{ij} based on standardized test scores (T_i). Let $\widehat{Y}_{ij} = f(s_{ij}(T_i))$ denote the applicant-program specific prediction of Y^*_{ij} constructed using the application score to program j, s_{ij} and a strictly increasing function whose roles is to relate units of Y_{ij} and application scores.[1] Programs define a selection threshold, τ_j, that is homogeneous to all applicants and revealed after they make their applications.[2] Let A_{ij} be an indicator

[1]As described in Section 3.3, applicants in Chile are endowed with a vector of test scores, $T_i = \left(T_i^1, ..., T_i^K\right)$, that is aggregated into a program-specific application score using known program-specific weights, $\alpha_j = (\alpha_j^1, ..., \alpha_j^K)$, with $\sum_k \alpha_j^k = 1$, such that $s_{ij} = \alpha_j' T_i$.

[2]In Chile, each program defines the number of openings that, given applications and the solution of a deferred-acceptance algorithm, determine selection thresholds. Our analysis does not require modeling the application decision.

function that takes value 1 if applicant i is accepted in program j, which happens if $s_{ij} \geq \tau_j$, i.e., $A_{ij} = 1\{s_{ij} \geq \tau_j\}$.

Bias in admissions may arise if the predictive ability of the test score varies between groups. Let $Y^*_{ij} = \widehat{Y}_{ij} + \epsilon_{ij}$, with ϵ_{ij} the prediction error. This implies that

$$\begin{aligned} A_{ij} = & \quad 1\left\{f^{-1}(\widehat{Y}_{ij}) \geq \tau_j\right\} & (3.1)\\ = & \quad 1\left\{f^{-1}(Y^*_{ij} - \epsilon_{ij}) \geq \tau_j\right\} & (3.2)\\ = & \quad 1\left\{Y^*_{ij} \geq f(\tau_j) + \epsilon_{ij}\right\} & (3.3)\\ \equiv & \quad 1\left\{Y^*_{ij} \geq h_{ij}\right\}, & (3.4) \end{aligned}$$

where $h_{ij} = f(\tau_j) + \epsilon_{ij}$ is the *effective threshold*. h_{ij} determines whether applicant i is accepted in program j given her latent outcome, Y^*_{ij}. Whenever ϵ_{ij} is positive, the test score underestimates Y^*_{ij} and, therefore, the applicant is set to a higher standard to be accepted in program j. Two applicants i_1 and i_2 with $Y^*_{i_1 j} = Y^*_{i_2 j}$ will be accepted at different rates in program j if $\epsilon_{i_1 j} \neq \epsilon_{i_2 j}$.

While prediction errors are ubiquitous to every prediction instrument, we say the test score is biased if there are systematic differences in prediction errors by demographic group that lead to systematic differences in effective thresholds and, therefore, to systematic disparities in college admissions. Let $G_i \in \{F, M\}$ denote the gender of applicant i. Bias, $\mathbf{B}$, is defined as

$$\mathbf{B} \;=\; \mathbb{E}[h_{ij}|G_i = F] - \mathbb{E}[h_{ij}|G_i = M], \qquad (3.5)$$

where the expectation integrates over applicants and programs. Positive values of $\mathbf{B}$ imply that the test score is biased against female applicants since, on average, they are set to stricter standards for acceptance into college conditional on latent outcomes.[3]

Identification Assume that for an applicant i that is enrolled in program j, Y^*_{ij} is (ex-post) observed. Note that application scores (s_{ij}) and nominal thresholds (τ_j) are also observed. Therefore, for the marginally selected applicants, that is, applicants for which $s_{ij} = \tau_j$, we have that $Y^*_{ij} = h_{ij}$. Using this identity, we can identify systematic differences in h_{ij} by gender by computing average differences in Y^*_{ij} between female and male marginal applicants. That is, differences in Y^*_{ij} at the

[3]In the psychometrics literature, the exercise of computing average prediction errors by a group is referred to as *differential prediction*. A related analysis, referred to as *differential validity*, consists of focusing on the R^2 of the prediction regardless of the bias. Our analysis focuses on the first diagnostic of test scores' predictive performance while being silent on the latter.

margin of selection identify **B**. This is a particular case of the more general families of marginal outcome tests used to estimate bias in usually more complex decision processes.[4]

Since nominal thresholds and latent outcomes are observed, we can in principle estimate the average prediction error by gender for the whole distribution of test scores by fitting regressions of observed outcomes on test scores complemented with proper selection corrections. In this paper, however, we focus on the margin rather than the whole distribution because our focus is on testing whether gender disparities in test scores' predictive accuracy led to gender gaps in college admissions. Focusing on the margin helps this purpose because the effects around the margin are more likely to have an impact on the application result. Differences in outcomes away from the margin will not necessarily identify differences in effective thresholds if treatment effects are heterogeneous, thus compromising the identification of **B**. To illustrate this point, suppose that prediction error disparities are concentrated in a segment of the distribution where no student is at the margin. Then, those disparities are unlikely to affect the allocation of applicants to programs. By contrast, if those differences are concentrated in segments of the distribution that many students qualify as marginals, then small gaps in predictive accuracy may lead to massive effects in allocations[5].

Specifying Y_{ij}^* The building block of marginal outcome tests is the existence of a well-defined outcome, Y_{ij}^*, that guides the selection process. In the context of college admissions, the definition of Y_{ij}^* is not trivial because the normative objectives of the tertiary education system are, most likely, multidimensional. College application processes usually deal with this by combining standardized tests with other instruments such as affirmative action or top percent policies.

In this paper, we focus on college admissions that are exclusively determined by test scores. We assume that the part of the admission system based on standardized tests seeks to assign applicants to programs based on some abstract notion of latent performance, possibly mediated by a combination of ability and match effects. Our empirical exercise uses two different variables to proxy Y_{ij}^*: first-year GPA and graduation rates. Conditional on specifying Y_{ij}^*, an additional assumption needed for identification is that Y_{ij}^* is not permeated by some bias similar in spirit to (3.5). If, for example, there are systematic differences in first-year GPA between men and

[4]Formal proofs for a more general version of this result can be found in Hull and Grau and Vergara. Outcome tests can be more complex since both predictions and selection thresholds are usually not observed.

[5]The related literature refer to this as the inframarginality problem (Knowles, Persico, and Todd; Anwar and Fang; Simoiu, Corbett-Davies, and Goel; Arnold, Dobbie, and Yang).

female conditional on latent performance because of some form of discrimination, then the outcome test will not identify **B**. We come back to this when discussing our results in Section 3.4 and 3.5.

Interpreting B Positive values of **B** imply that the application process is biased against women. How should we interpret this conclusion? Since nominal selection thresholds and test scores-based predictions are gender neutral by design, we can rule out canonical statistical and taste-based discrimination and assert that **B** exclusively captures gender gaps in the tests' predictive ability.[6] In addition, since the process is carried out by a presumably gender-neutral algorithm rather than human decision-makers, the estimated bias can accommodate notions of institutional discrimination discussed in Bohren, Hull, and Imas.

Marginal outcome tests, however, do not inform about the structural forces that explain the differential predictive ability of the standardized test. **B** is a reduced-form diagnostic that can accommodate different narratives. The design of the test could hurt female students for reasons unrelated to potential performance (Niederle and Vesterlund; Coffman and Klinowski; Saygin and Atwater) or gender-specific teacher fixed-effects could affect long-run test-taking performance (Dee; Carrell, Page, and West; Paredes; Muralidharan and Sheth; Lim and Meer; Martinez). Our analysis cannot disentangle between these – or many other – potential sources of bias.

3.3 Institutional setting and data

The centralized college admission process in Chile fully relies on standardized test scores and high school GPA. For our empirical exercises, we use college admission (between 2012 and 2017 years) and college records (enrollment and graduation information from 2012 to 2022) from administrative data. The gender gap in students' performance for 1st-year GPA is 0.35 - equivalent to 0.4 standard errors- and significant, and for graduation is 0.12 percentage points and significant.

Institutional setting

The Chilean college system consists of a group of older and more selective universities with a centralized admission process and a group of newer institutions with

[6]In settings where the selection process is carried out by heterogeneous decision-makers, disentangling between drivers of discrimination –accurate statistical discrimination, biased predictions, and taste-based discrimination– is not possible without additional structure (Arnold, Dobbie, and Yang; Bohren et al.; Arnold, Dobbie, and Hull; Hull).

particular admission systems. In this article, we focus on the first group. The selection in the centralized admission system is based solely on standardized admission test scores and high school GPA [7].

Before the end of the school year institutions provide public information about the number of vacancies for each program available in the next application process and the weights associated with test scores and high school GPA to calculate applicants' application scores.

Standardized tests are taken at the end of the school year by high school graduates. Test scores are observed by applicants and then students submit a list with the ranking of up to ten programs (college-major combinations) based on their preferences.

A deferred acceptance (DA) algorithm is used to assign students to programs based on their reported preference and their application scores. The DA algorithm generates offers (at most one per student) to the most preferred option that the student is eligible for, based on their program-specific application score.

Data

We use college administrative data provided by the Department of Educational Evaluation, Measurement and Registration (DEMRE for their acronym in Spanish), the Higher Education Information Service (SIES) at the Ministry of Education, and the Council of Presidents of Chilean Universities (CRUNCH for their acronym in Spanish). The records are at the student level and include information about sociodemographics, high-school graduation, admission test results, admission ranking, program preferences, and performance measures.

We estimate our results for slightly different entrance cohorts depending on the outcome considered - first-year GPA and graduation probabilities.

For the first-year GPA outcome, we observe in our data all the students who enrolled in a program in a selective university between the years 2013 and 2017. This data was obtained through an agreement with the Council of Presidents of Chilean Universities and it only contains information on the year, enrolled program, application scores, sex of the student, and first-year GPA. Because this data is not part of the publicly available data, the identification number in this sample doesn't allow for matches with publicly available data on college students and we are limited to the data directly provided.

[7] There are between 33 and 36 universities between 2012 and 2017 offering around 1,400 programs in total.

For the college performance measure based on graduation, we look at the universe of students enrolled in a selective college in 2012 and 2013. This is publicly available data. By merging at the student level the graduation records between 2013 and 2022, we construct an indicator variable that takes the value of one if the student ever graduates from college. Most programs in Chile have a nominal duration between 5 to 7 years, however, most students take between one and three extra years to finish it. Therefore our focus is on cohorts with enough time to observe graduation [8]. At an individual level, we incorporate socioeconomic characteristics, application test scores, program application preferences, and high school characteristics from publicly available data [9].

In both cases, the sample used in analysis to calculate the gap between female and male performance is restricted to programs with at least one man and one woman in the 10th percentile around the admission cutoff.

Descriptive statistics

In our cohorts, there are around 100,000 students that apply and around 70,000 students that enroll in a selective college each year. Each year students apply on average to 5 programs (with a minimum of 1 and a maximum of 10).

Tables 3.1 and 3.2 present the main characteristics of our two samples, 2013 - 2017 and 2012-2013 without restrictions, restricted to programs with at least one man and one woman in the 10th percentile around the admission cutoff, and only for students in the 10th percentile around the admission cutoff enrolled in programs with at least one man and one woman in the 10th percentile around the admission cutoff.

The outcome row data presents some preliminary evidence in favor of the hypothesis of gender bias in admission.

3.4 Empirical analysis

The outcome test estimation shows that on average women at the margin of being admitted have consistently higher 1st-year GPAs in college - 0.137 points- and higher probability of graduation - 7 percentage points. We present a simple counterfactual exercise to quantify the bias in terms of college applications, taken as

[8] We cannot use earlier application cohorts because in 2011 the universities included in the central admission system changed.

[9] Most of the education publicly available data can be found in `https://datosabiertos.mineduc.cl/`

given the outcome test results. In these cases, trying to close the estimated latent performance gap could improve the outcome of admission for around 2% of the female applicants.

Outcome test

Empirical strategy Recall from Section 3.2 that the bias diagnostic, $\mathbf{B}$, is identified by differences in means within samples of marginally selected applicants. We implement this by estimating the following regression by OLS:

$$Y_i = \alpha_0 + \alpha_1 \cdot F_i + \gamma_{j(i)} + \varepsilon_i, \tag{3.6}$$

where Y_i is an outcome (first-year GPA or graduation indicator), F_i is an indicator variable that takes value 1 if applicant i is female, $j(i)$ is the program that student i is enrolled into, $\gamma_{j(i)}$ is a program fixed-effect that makes grades and graduation rates comparable across programs, and ε_i is the projection error. The main specification restricts the sample to programs with at least one man and one woman in the 10th percentile around the admission cutoff, and to students in the 10th percentile around the admission cutoff, so α_1 identifies $\mathbf{B}$.

Main results Table 3.3 presents the main results for our outcome test. With our preferred specification in columns (2) and (4) we see that there is a positive and significant gap between female and male college performance measured by their 1st-year GPA or their graduation rates. In both cases, the gap decreases when program fixed effects are included, suggesting that on average women enroll in programs where the marginal students have higher 1st-year GPAs and higher graduation rates. The respective performance gaps between females and males are 0.14 1st-year GPA points and 7 percentage points for graduation.

Heterogeneities We explore heterogeneous patterns in the estimated bias. Results are presented in Figure 3.1, when the outcome considered, is 1st-year GPA, and 3.2 when the outcome is graduation probability.

First, we estimate equation (3.6) by quartile of cutoff scores, the proportion of females enrolled in the program, and the proportion in the application score formula assigned to high school GPA components. For 1st-year GPA we observe some heterogeneous patterns in cutoff scores and the grade components. In less selective programs, therefore, for students with lower application scores, we observe higher gaps in 1st-year performance between females and males. However, in the highest cutoff quartile, the gap is still positive but significantly smaller. For the quartile with higher program enrollment of females, there is some evidence suggesting that the bias is bigger. Finally, the gap observed in programs with a small weight in the high school GPA application score formula or high weight associated with standardized test scores presents a significantly higher gap in 1st-year performance. The reverse is true for programs in the last quartile; however, even when high school GPA components are more than 45%, the gap in performance remains positive.

The longer-term performance outcome, the graduation indicator, doesn't present any distinctive pattern through the quartiles.

Gender gaps in college admissions

How large is the estimated bias? We answer this question by transforming outcome units to test score points to simulate alternative college allocations that adjust for the bias.

From outcomes to test score points To transform outcome units to test score points, we estimate the following linear regression:

$$\widehat{Y_i} \;=\; \beta_0 + \beta_1 \cdot Y_i + v_i, \tag{3.7}$$

where Y_i accounts for the outcomes considered in the previous subsection and $\widehat{Y_i}$ is the application test-score. We use the estimates of β_1 to transform differences in effective thresholds measured in outcomes' units into test score units.

For the graduation outcome, we use the predicted probability for graduation in Y_i. The justification is that given that the graduation outcome is distant in the future the correlation between the standardized test score is weak and therefore the transformation suggests a small equivalence to application points but not due to a small gap. In other words, there are several shocks that affect the final indicator that may be not related at all to the standardized test score outcome so using the probability of graduation we get a cleaner variable that is more sensitive in the first year.

This predicted probability is estimated with a simple linear probability model that considers the sex of the student and indicators if the student has a wage job, lives in an urban area, has access to private health insurance, and their parents graduate from college [10].

Tables 3.4 and 3.6 show the results. In both cases, estimated from the 1st year college GPA and from the probability of graduation, the results are consistent with

[10] In appendix Table C.1 we present the results for Y_i as the graduation indicator directly. In this case, the application score equivalence is 2 application points.

women facing a selection threshold that is, on average, three points larger. Even though the number may appear as small when considering that the distribution of the application score has a mean of 500 and a standard deviation of 110 points, we assess the consequences of this gap in the next section with three different counterfactual exercises.

Counterfactual exercises Our first counterfactual exercise simulates the change in enrollment observed by men if they were admitted based on the effective threshold observed by women. This exercise is a partial equilibrium result in which we count the number (or the proportion) of male students that wouldn't make the admission cutoff when it becomes stricter. In Figure 3.3 we present the distribution of programs observed between 2013 and 2017 based on the change in the admission of males. The average change of student admissions for 2013 is 4.3% on total enrollment and around 50% of the program loses between 1 and 7 male students when exposing men to women's effective threshold.

Our second set of counterfactual exercises simulates the change admission offers in the entire system if, in order to close the estimated effective gap, (1) female students were to experience a 3 points adjustment in their application scores or (2) programs were requested higher percentages in the admission score formula associated to the GPA components. In order to implement this exercise we only consider 2013 applicants. Using their entire rank order list of preferences and a deferred acceptance algorithm that replicates 99.9% of the observed admissions we replicate the admission offer by running the assignment algorithm with the adjusted application scores. The change in the application scores is straightforward for the first case - we just add 3 points to each women's application score.

The second set of general equilibrium counterfactuals is justified by the previously observed heterogeneity. Programs with less weight in the standardized test score present smaller gaps in 1st-year performance, therefore, implementing this reduction should help to close the observed gap. In this case, we increase the high school GPA component for how much is needed to get to the minimum considered in the exercise. The same adjustment is subtracted from the standardized test score component in such a way that the proportion used for each test score is not altered [11].

Table 3.5 present the results from the general equilibrium counterfactual exercises. In both cases around 2% of female applicants would be better off with the proposed changes. Note that the majority of the improvement comes from women that are

[11] For example, if we consider a minimum of 30% in the GPA component but the program assigns only 20% to it we add to the application score 10% $\times$ GPA score and we subtract 10% $\times$ weighted average of standardized test scores components

able to reach more preferred programs and not from women getting admitted in this new scenario that would have not had an offer with the adjustment. Given that women are more likely than men to list programs from multiple fields and the fact that STEM programs are on average more selective than other field programs, this evidence suggests that the effort of closing the effective selective threshold could have an impact later on in the labor market through changing the programs that women get access to.

Moreover, it is clear that doing the 3 points adjustment improves female admissions by imposing a negative change in admission for males; however, when increasing the % of the high school GPA components there is more "distortion" created into the entire system, with a significant amount of women with their admission affected negatively [12]. All things considered, the consequences of the differential predictive ability are not negligible in terms of misallocation.

3.5 Discussion

This paper uses marginal outcome tests and administrative data from Chile to explore whether standardized tests for college admissions are biased against female applicants. Using different proxies for college performance, we find that standardized tests systematically under-predict women's outcomes. We express the bias estimates as differences in effective selection thresholds which we suggest affect approximately 2% of female applicants. This result, even though marginal, is equivalent in magnitude to some of the affirmative action and diversity initiatives implemented in the system. Moreover, noting that the estimated bias is smaller in programs that put more weight on high-school outcomes for the application scores, we estimate that increasing the mandatory minimum associated with the high school GPA components can lead to similar net admission changes for females - improvement in the outcome for around 2% of them - however, this policy has more people with their admission affected.

One potential concern when interpreting the gap as a gender bias in college admission is that if females at the margin have better socioeconomic conditions, then the observed gap performance could be explained by the different types of shocks that male and female students are exposed to due to while in college and not due to a difference in their latent performance. In Table 3.7 we present the results from Equation 3.6 using as dependent variable different socioeconomic characteristics of students. The results are contrary to the raised concern; if anything, women at

[12]Note that the positive net effect for women is observed only when the minimum GPA weight is at least 40%.

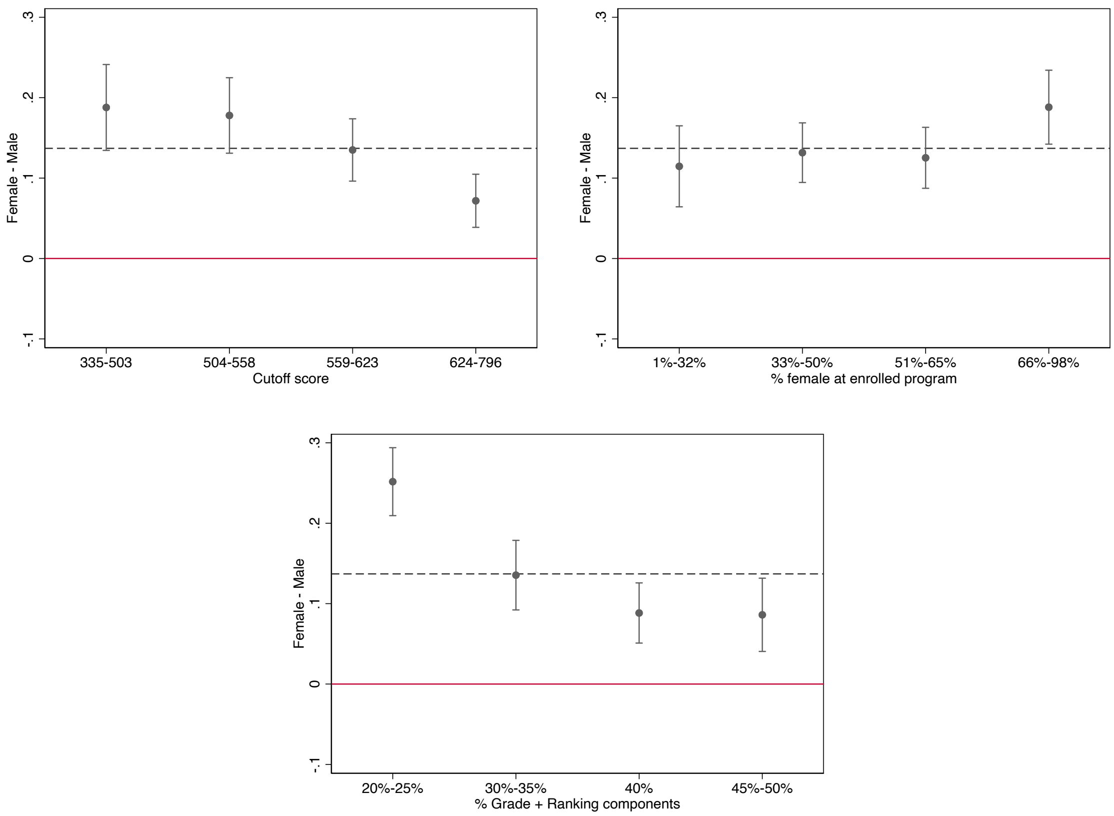
the margin are from more disadvantaged backgrounds than men, and, therefore, the results could be interpreted as a lower bound.

As we acknowledge in Section 3.2, our estimations are a reduced-form diagnostic of bias, similar in spirit to a “sufficient statistic”. This implies that, while our estimates of aggregate bias are robust to alternative structural models, we cannot inform about the latter and, therefore, our results are silent about the first-order policy recommendations to address the sources of the bias. However, as illustrated by the counterfactual exercises, second-best policy avenues can be designed based on our analysis. Our analysis suggests that gender-specific score inflation and changes in the weight given to high-school outcomes in the application scores can work as short-term policies to deal with the estimated bias. Gender-specific score inflation could be phased out if the structural sources of bias are addressed, for example, by recomputing the correction factor using moving averages.

Our intuition is that the proposed short-term policies are likely to work in the correct direction, however, we call for caution since more research is needed to understand its optimal design and other potential unintended consequences. For example, the counterfactual analysis takes both high-school outcomes and application strategies as given, but they both could be affected by the aforementioned policies. Recent evidence suggests there are gender-specific aspects in optimal application strategies (Turner and Bowen; Bordón, Canals, and Mizala). It is possible that application behavior could be affected by policies that affect selection at the margin. In addition, giving a more important role to high-school outcomes could affect high-school practices, for example, creating incentives for grade inflation (Fajnzylber, Lara, and León).

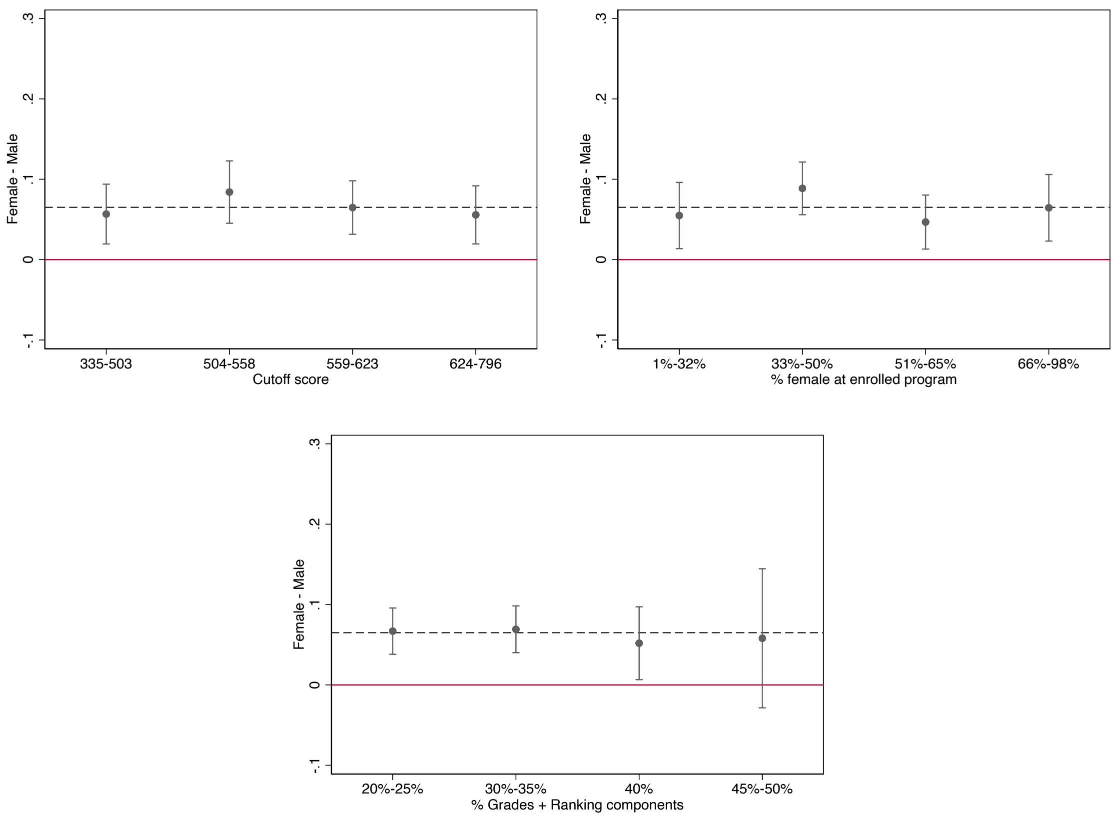
3.6 Figures and tables

Figure 3.1: 1st year GPA differences at enrolled program



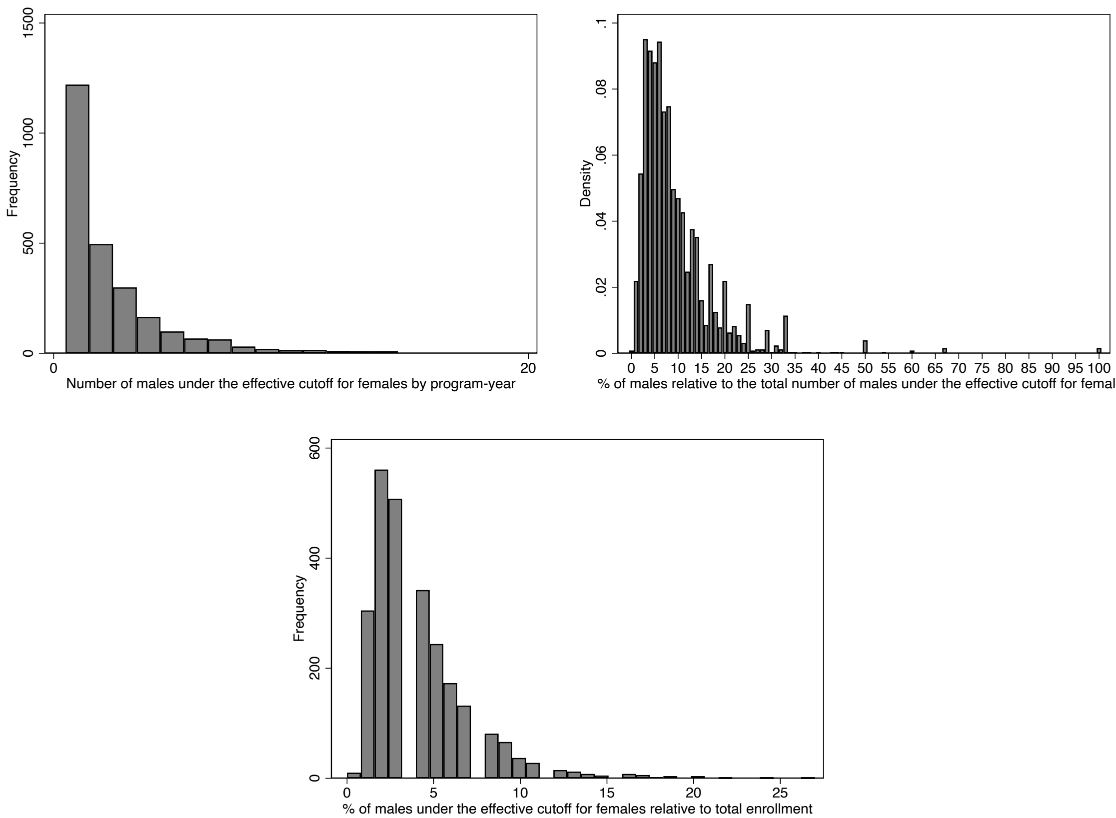
Notes: We focus on students enrolled between 2013 and 2017. The effect is estimated for students around the cutoffs - 10% of enrolled students around it. Estimates only consider programs with at least one male and one female at the margin. For our main specification, in each quartile, we include year and program fixed effects.

Figure 3.2: Graduation rates differences at enrolled program



Notes: We focus on students enrolled between 2012 and 2013. The effect is estimated for students around the cutoffs - 10% of enrolled students around it. Estimates only consider programs with at least one male and one female at the margin. For our main specification, in each quartile, we include year and program fixed effects.

Figure 3.3: Partial simulation of changes in male admission by program - year



Notes: in each case, we have the distribution of program-year between 2013 and 2017 based on (1) the number of males not admitted in each program if they have to face the female effective threshold (3 points less); (2) the proportion of males not admitted in each program with the female effective threshold relative to the total number of males in each program; and (3) the proportion of males not admitted in each program with the female effective threshold relative to total enrollment in each program.

Table 3.1: Description sample 2013 - 2017

	Margin w/restriction		**Total w/restriction**		**Total wo/restriction**	
	Female	Male	Female	Male	Female	Male
Outcomes						
Grades 1st year	4.56	4.27	4.84	4.57	4.87	4.52
(Std. Dev.)	(0.94)	(1.01)	(0.89)	(0.97)	(0.90)	(1.00)
Program Charact.						
Cutoff score	578.3	576.7	575.3	578.4	566.4	566.7
(Std. Dev.)	(76.8)	(79.8)	(76.7)	(80.3)	(75.2)	(79.4)
% Female	55.6	43.8	57.3	41.7	61.2	38.4
(Std. Dev.)	(18.9)	(18.7)	(18.3)	(17.9)	(21.4)	(20.5)
% Grades + Ranking	35.9	35.4	36.1	35.2	36.4	36.2
(Std. Dev.)	(10.3)	(10.5)	(10.3)	(10.5)	(10.4)	(10.6)
Observations	11,776	12,896	124,982	127,947	166,570	168,277

Notes: This table presents the mean and standard deviation of the main characteristics available for the sample of enrolled students between 2013 - 2017. The first two columns present the characteristics of the sample of students in the 10th percentile around the admission cutoff in programs with at least one man and one woman in the 10th percentile around the admission cutoff. The 3rd and 4th columns present the characteristics of the sample of students in programs with at least one man and one woman in the 10th percentile around the admission cutoff. And the last two columns present the characteristics of all the students enrolled.

Table 3.2: Description sample 2012 - 2013

	Margin w/restriction		**Total w/restriction**		**Total wo/restriction**	
	Female	Male	Female	Male	Female	Male
Outcomes						
Graduation	0.45	0.36	0.54	0.45	0.54	0.43
(Std. Dev.)	(0.50)	(0.48)	(0.50)	(0.50)	(0.50)	(0.49)
GPA score	570.57	545.06	622.80	596.65	616.21	588.17
(Std. Dev.)	(97.51)	(99.12)	(88.57)	(94.93)	(89.58)	(95.28)
Verbal score	569.1	568.5	600.3	601.1	594.3	591.5
(Std. Dev.)	(77.0)	(78.8)	(77.7)	(79.0)	(76.8)	(79.6)
Math score	564.4	589.1	591.0	619.8	584.5	610.7
(Std. Dev.)	(73.7)	(80.5)	(73.2)	(81.8)	(72.4)	(80.6)
History score	324.4	325.6	345.0	340.4	341.7	330.3
(Std. Dev.)	(281.3)	(294.7)	(294.4)	(311.1)	(290.2)	(306.7)
Science score	362.5	393.8	384.9	429.2	379.6	427.1
(Std. Dev.)	(261.6)	(266.5)	(274.9)	(274.4)	(272.9)	(267.5)
Moth w/college	0.2	0.2	0.2	0.2	0.2	0.2
(Std. Dev.)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)	(0.4)
Father w/college	0.3	0.3	0.3	0.3	0.3	0.3
(Std. Dev.)	(0.4)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
Private health	0.4	0.4	0.4	0.4	0.4	0.4
(Std. Dev.)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
N applications	5.3	5.3	5.1	5.0	5.1	5.0
(Std. Dev.)	(2.5)	(2.5)	(2.4)	(2.4)	(2.4)	(2.4)
Ranking offer	2.3	2.2	1.8	1.8	1.8	1.7
(Std. Dev.)	(1.5)	(1.5)	(1.3)	(1.2)	(1.3)	(1.2)
Program Charact.						
Cutoff score	576.7	579.3	574.8	580.3	566.6	567.6
(Std. Dev.)	(71.6)	(77.4)	(71.1)	(77.8)	(71.2)	(78.2)
% Female	55.2	42.2	56.6	40.7	60.3	37.7
(Std. Dev.)	(19.4)	(18.4)	(18.8)	(17.8)	(21.6)	(20.3)
% Grades + Ranking	29.3	29.5	29.4	29.4	29.7	30.2
(Std. Dev.)	(8.0)	(8.2)	(8.1)	(8.1)	(8.1)	(8.4)
Observations	4,882	5,615	54,360	58,070	71,426	75,238

Notes: This table presents the mean and standard deviation of the main characteristics available for the sample of enrolled students between 2012 - 2013. The first two columns present the characteristics of the sample of students in the 10th percentile around the admission cutoff in programs with at least one man and one woman in the 10th percentile around the admission cutoff. The 3rd and 4th columns present the characteristics of the sample of students in programs with at least one man and one woman in the 10th percentile around the admission cutoff. And the last two columns present the characteristics of all the students enrolled.

Table 3.3: Differences in outcomes between female and males

	1st-year GPA		Graduation	
	(1) without FE	(2) with FE	(3) without FE	(4) with FE
Female	0.283***	0.137***	0.086***	0.065***
	(0.012)	(0.011)	(0.010)	(0.009)
Mean of Dep Variable	4.408	4.408	0.402	0.402
Program Fixed Effects		✓		✓
Observations	24,672	24,672	10,497	10,497

Notes: This table shows the differences in outcomes between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin. Columns 1 and 2 present the difference in 1st year GPA calculated without and with year and program fixed effects, respectively. The sample considers students enrolled in 1st year of college between 2013 and 2017. Columns 3 and 4 present the difference in graduation rate, also without and with year and program fixed effects, respectively. This sample considers students enrolled in 1st year of college between 2012 and 2013.

Table 3.4: Translation of gender differences in GPA into applications score points

Application score equivalence to grade difference		
	(1) 1st-year GPA	(2) Application score
Female	0.137***	
	(0.011)	
1st year GPA		22.777***
		(0.142)
Constant	4.343***	517.027***
	(0.007)	(0.683)
Program Fixed Effect	Yes	No
Cutoff margin	Yes	No
Observations	24,672	248,323
Application score equivalence	3.11	

Notes: Column 1 of the table shows the difference in 1st year GPA between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application score and 1st year GPA for all the students enrolled between 2013 and 2017 in programs with at least one male and one female at the margin. We use this relationship to translate the differences in 1st year GPA into differences in application scores presented at the bottom of the table.

Table 3.5: Translation of gender differences in graduation rates into applications score points

Application score equivalence to graduation difference

	(1) Graduation	(2) Score
Female	0.065***	
	(0.009)	
Pred. Grad		47.175***
		(3.846)
Constant	0.372***	591.808***
	(0.006)	(1.884)
Program Fixed Effect	Yes	No
Cutoff margin	Yes	No
Observations	10,497	92,192
Application score equivalence	3.06	

Notes: Column 1 of the table shows the difference in graduation rates between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application scores and students' predicted graduation probability based on a simple linear probability model (details in Table C.2) that takes gender, and indicator variables for: a student wage job, living in an urban area, having private health, his/her mother or father having a college degree. We use this relationship to translate the differences in graduation rates into differences in application scores presented at the bottom of the table.

Table 3.6: Summary of the total number of affected students under counterfactual scenarios

	3 point adjustment		**30% GPA**		**40% GPA**		**50% GPA**	
	(1) Female	(2) Male	(3) Female	(4) Male	(5) Female	(6) Male	(7) Female	(8) Male
Positive change								
More preferred program	852	21	2,376	2,423	3,849	2,777	3,849	2,783
Some program	223	9	663	551	972	595	971	593
Better outcome	1,075	30	3,039	2,974	4,821	3,372	4,820	3,376
Negative change								
Less preferred program	36	940	2,632	2,358	2,168	2,602	2,160	2,604
No program	17	246	720	638	945	1,033	947	1,033
Worst outcome	53	1,186	3,352	2,996	3,113	3,635	3,107	3,637

Notes: Admission simulations are run on the cohort of 2013 applicants (107,546 students). Columns 1 and 2 present the number of female and male applicants with their admission affected when 3 points are added to the female application score. Columns 3 and 4, 5 and 6, and 7 and 8, present the number of female and male applicants with their admission affected when the GPA components are forced to have a minimum of 30%, 40%, and 50%, respectively. In the three cases, the adjustment keeps the proportions assigned to each test score component and increases the GPA component to get to the minimum established. For example, if the minimum is 30% and before the GPA + ranking component was at 20%, 10% extra weight is added to the GPA component, and 10% is subtracted from the weighted average of test score components.

Table 3.7: Differences in family characteristics between female and males

	(1) Graduation	(2) Private health	(3) Mother with college	(4) Father with college
Female	0.065***	-0.018*	-0.014	-0.003
	(0.009)	(0.009)	(0.009)	(0.010)
Mean of Dep Variable	0.402	0.388	0.222	0.289
Program Fixed Effects	✓	✓	✓	✓
Observations	10,497	10,497	9,751	9,050

Notes: This table shows the differences in family characteristics between females and males around the cutoff - 10% of students - considering the sample of students enrolled between 2012 and 2013 in 1st year of college, in programs with at least one male and one female in that margin, and with year and program fixed effect. Column 1 presents the difference in graduation rate; column 2 presents the difference in the probability of having private health insurance; and columns 3 and 4 present the difference in the probability of having a mother or a father, respectively, with a college degree.

Table .8: Translation of gender differences in graduation into applications score points

Application score equivalence to graduation difference

	(1) Graduation	(2) Application score
Female	0.065***	
	(0.009)	
Graduation		29.671***
		(0.394)
Constant	0.372***	600.737***
	(0.006)	(0.277)
Program Fixed Effect	Yes	No
Cutoff margin	Yes	No
Observations	10497	112430
Application score equivalence	1.93	

Notes: Column 1 of the table shows the difference in graduation rates between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application scores and graduation for all the students enrolled between 2012 and 2013 in programs with at least one male and one female at the margin. We use this relationship to translate the differences in graduation into differences in application scores presented at the bottom of the table.

Table .9: Translation of gender differences in GPA into applications score points

	(1) Grad
Female	0.075***
	(0.003)
Wage job	-0.047***
	(0.006)
Urban housing	0.035***
	(0.003)
Private health	0.020***
	(0.003)
Mother with college	0.001
	(0.004)
Father with college	0.003
	(0.004)
Program Fixed Effect	Yes
Cutoff margin	No
Observations	120,503

Notes: Column 1 of the table shows the differences in 1st year GPA between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application score and 1st year GPA for all the students enrolled between 2013 and 2017 in programs with at least one male and one female at the margin. We use this relationship to translate the differences in 1st year GPA into differences in application scores presented at the bottom of the table.

Bibliography

[1] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match". In: *American Economic Review* 99.5 (2009), 1954–78.

[2] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. "The new york city high school match". In: *American Economic Review* 95.2 (2005), 364–367.

[3] Atila Abdulkadiroğlu and Tayfun Sönmez. "School choice: A mechanism design approach". In: *American economic review* 93.3 (2003), 729–747.

[4] Atila Abdulkadiroğlu et al. "Do parents value school effectiveness?" In: *American Economic Review* 110.5 (2020), 1502–39.

[5] Nikhil Agarwal, Charles Hodgson, and Paulo Somaini. *Choices and outcomes in assignment mechanisms: The allocation of deceased donor kidneys*. Tech. rep. National Bureau of Economic Research, 2020.

[6] Josefa Aguirre, Juan Matta, and Ana Marıa Montoya. "Joining the Men's Club: The Returns to Pursuing High-earnings Male-dominated Fields for Women". In: *Working Paper* (2020).

[7] Josefa Aguirre, Juan Matta, and Ana Marıa Montoya. "The Long-term Effects of College Peers". In: *Working Paper* (2021).

[8] Joseph G Altonji and Rebecca M Blank. "Race and gender in the labor market". In: *Handbook of Labor Economics* 3 (1999), 3143–3259.

[9] Massimo Anelli. "The returns to elite university education: A quasi-experimental analysis". In: *Journal of the European Economic Association* 18.6 (2020), 2824–2868.

[10] Julia Angwin et al. "Machine bias". In: *ProPublica, May* 23.2016 (2016), 139–159.

[11] S. Anwar and H. Fang. "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence". In: *American Economic Review* 96.1 (2006), 127–151.

[12] Peter Arcidiacono, Esteban M Aucejo, and V Joseph Hotz. "University differences in the graduation of minorities in STEM fields: Evidence from California". In: *American Economic Review* 106.3 (2016), 525–62.

[13] Peter Arcidiacono and Michael Lovenheim. "Affirmative action and the quality-fit trade-off". In: *Journal of Economic Literature* 54.1 (2016), 3–51.

[14] Peter Arcidiacono et al. "Affirmative action and university fit: Evidence from Proposition 209". In: *IZA Journal of Labor Economics* 3.1 (2014), 1–29.

[15] Peter Arcidiacono et al. "Does affirmative action lead to mismatch? A new test and evidence". In: *Quantitative Economics* 2.3 (2011), 303–333.

[16] D. Arnold, W. Dobbie, and P. Hull. "Measuring racial discrimination in bail decisions". In: *Working Paper* (2020).

[17] D. Arnold, W. Dobbie, and C. Yang. "Racial bias in bail decisions". In: *Quarterly Journal of Economics* 133.4 (2018), 1885–1932.

[18] David Arnold, Will Dobbie, and Peter Hull. "Measuring racial discrimination in algorithms". In: *AEA Papers and Proceedings* 111 (2021), 49–54.

[19] David H Autor. "Skills, education, and the rise of earnings inequality among the "other 99 percent"". In: *Science* 344.6186 (2014), 843–851.

[20] Surendrakumar Bagde, Dennis Epple, and Lowell Taylor. "Does affirmative action work? Caste, gender, college quality, and academic success in India". In: *American Economic Review* 106.6 (2016), 1495–1521.

[21] Andrew C Baker, David F Larcker, and Charles CY Wang. "How much should we trust staggered difference-in-differences estimates?" In: *Journal of Financial Economics* 144.2 (2022), 370–395.

[22] Alan Benson, Danielle Li, and Kelly Shue. ""Potential" and the Gender Promotion Gap". In: *Working Paper* (2021).

[23] Richard Berk et al. "Fairness in criminal justice risk assessments: The state of the art". In: *Sociological Methods & Research* 50.1 (2021), 3–44.

[24] Graziella Bertocchi and Monica Bozzano. *Gender gaps in education*. Springer, 2020.

[25] Marianne Bertrand. "New perspectives on gender". In: *Handbook of Labor Economics*. Vol. 4. Elsevier, 2011, 1543–1590.

[26] Dan A Black and Jeffrey A Smith. "How robust is the evidence on the effects of college quality? Evidence from matching". In: *Journal of econometrics* 121.1-2 (2004), 99–124.

[27] Sandra E Black, Kalena E Cortes, and Jane Arnold Lincove. "Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies". In: *American Economic Review* 105.5 (2015), 604–10.

[28] Sandra E Black, Kalena E Cortes, and Jane Arnold Lincove. "Efficacy Versus Equity: What Happens When States Tinker With College Admissions in a Race-Blind Era?" In: *Educational Evaluation and Policy Analysis* 38.2 (2016), 336–363.

[29] Sandra E Black, Jeffrey T Denning, and Jesse Rothstein. *Winners and Losers? The Effect of Gaining and Losing Access to Selective Colleges on Education and Labor Market Outcomes*. Tech. rep. National Bureau of Economic Research, 2020.

[30] Sandra E Black, Jeffrey T Denning, and Jesse Rothstein. "Winners and losers? The effect of gaining and losing access to selective colleges on education and labor market outcomes". In: *American Economic Journal: Applied Economics* (2021).

[31] R. Blanco, R. Hutt, and H. Rojas. "Reform to the Criminal Justice System in Chile: Evaluation and Challenges". In: *Loyola University Chicago International Law Review* 2.2 (2004), 253–269.

[32] Francine D Blau and Lawrence M Kahn. "The gender wage gap: Extent, trends, and explanations". In: *Journal of Economic Literature* 55.3 (2017), 789–865.

[33] Zachary Bleemer. "Affirmative action, mismatch, and economic mobility after California's Proposition 209". In: *Working Paper* (2021).

[34] Zachary Bleemer. "Affirmative action, mismatch, and economic mobility after California's Proposition 209". In: *The Quarterly Journal of Economics* 137.1 (2022), 115–160.

[35] Zachary Bleemer. "Top Percent Policies and the Return to Postsecondary Selectivity, by Zachary Bleemer, CSHE 1.21". In: (2021).

[36] Zachary Bleemer and Aashish Mehta. "Will studying economics make you rich? a regression discontinuity analysis of the returns to college major". In: *American Economic Journal: Applied Economics* (2020).

[37] J Aislinn Bohren, Peter Hull, and Alex Imas. "Systemic discrimination: Theory and measurement". In: *Working Paper* (2023).

[38] J Aislinn Bohren et al. "Inaccurate statistical discrimination". In: *Working Paper* (2020).

[39] Paola Bordón, Catalina Canals, and Alejandra Mizala. "The gender gap in college major choice in Chile". In: *Economics of Education Review* 77 (2020), 102011.

[40] Kaja Høiseth Brugård and Falch Torberg. "Post-compulsory education and imprisonment". In: *Labour Economics* 97 (2013), 97–106.

[41] Claudia Buchmann, Thomas A DiPrete, and Anne McDaniel. "Gender inequalities in education". In: *Annu. Rev. Sociol* 34 (2008), 319–337.

[42] Julia Burdick-Will. "School violent crime and academic achievement in Chicago". In: *Sociology of education* 86.4 (2013), 343–361.

[43] Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. "Robust data-driven inference in the regression-discontinuity design". In: *Stata Journal* 14.4 (2014), 909–946.

[44] Sebastian Calonico, Matias D. Cattaneo, and Rocio Titiunik. "Robust Non-parametric Confidence Intervals for Regression-Discontinuity Designs". In: *Econometrica* 82 (2014), 2295–2326.

[45] Sebastian Calonico et al. "Regression Discontinuity Designs Using Covariates". In: *Econometrica* 101.3 (2019), 442–451.

[46] David Card. "The causal effect of education on earnings". In: *Handbook of Labor Economics* 3 (1999), 1801–1863.

[47] David Card and Laura Giuliano. "Universal screening increases the representation of low-income and minority students in gifted education". In: *Proceedings of the National Academy of Sciences* 113.48 (2016), 13678–13683.

[48] David Card and Alan B Krueger. "Would the elimination of affirmative action affect highly qualified minority applicants? Evidence from California and Texas". In: *ILR Review* 58.3 (2005), 416–434.

[49] David Card and Jesse Rothstein. "Racial segregation and the black–white test score gap". In: *Journal of Public Economics* 91.11-12 (2007), 2158–2184.

[50] David Card et al. "Gender differences in peer recognition by economists". In: *Working Paper* (2021).

[51] Scott E Carrell, Marianne E Page, and James E West. "Sex and science: How professor gender perpetuates the gender gap". In: *The Quarterly Journal of Economics* 125.3 (2010), 1101–1144.

[52] Matias D. Cattaneo, Michael Jansson, and Xinwei Ma. *Simple Local Polynomial Density Estimators*. Working paper. University of Michigan, 2017.

[53] Raj Chetty et al. "Income segregation and intergenerational mobility across colleges in the United States". In: *The Quarterly Journal of Economics* 135.3 (2020), 1567–1633.

[54] Raj Chetty et al. *Mobility report cards: The role of colleges in intergenerational mobility*. Tech. rep. national bureau of economic research, 2017.

[55] Katherine B Coffman and David Klinowski. "The impact of penalties for wrong answers on the gender gap in test scores". In: *Proceedings of the National Academy of Sciences* 117.16 (2020), 8794–8803.

[56] Sarah R Cohodes and Joshua S Goodman. "Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy". In: *American Economic Journal: Applied Economics* 6.4 (2014), 251–85.

[57] Carolina Concha-Arriagada. *Should I Stay, or Should I go? Strategic Responses to Improve College Admission Chances*. Tech. rep. Working paper, 2022.

[58] Philip J. Cook and Songman Kang. "Birthdays, Schooling, and Crime: Regression-Discontinuity Analysis of School Performance, Delinquency, Dropout, and Crime Initiation". In: *American Economic Journal: Applied Economics* 8.1 (2016), 33–57.

[59] Tomás Cortés, Nicolás Grau, and Jorge Rivera. *Juvenile Incarceration and Adult Recidivism*. Working Paper. Universidad de Chile, 2020.

[60] Jaime Couso and Julio Duce. *Juzgamiento penal de adolescentes*. LOM ediciones, 2013.

[61] Rachel Croson and Uri Gneezy. "Gender differences in preferences". In: *Journal of Economic literature* 47.2 (2009), 448–74.

[62] Julie Berry Cullen, Mark C Long, and Randall Reback. "Jockeying for position: Strategic high school choice under Texas' top ten percent plan". In: *Journal of Public Economics* 97 (2013), 32–48.

[63] Zoë B Cullen and Ricardo Perez-Truglia. "The old boys' club: Schmoozing and the gender gap". In: *Working Paper* (2021).

[64] Stacy B Dale and Alan B Krueger. "Estimating the effects of college characteristics over the career using administrative earnings data". In: *Journal of human resources* 49.2 (2014), 323–358.

[65] Stacy Berg Dale and Alan B Krueger. "Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables". In: *The Quarterly Journal of Economics* 117.4 (2002), 1491–1527.

[66] Clément De Chaisemartin and Xavier d'Haultfoeuille. "Two-way fixed effects estimators with heterogeneous treatment effects". In: *American Economic Review* 110.9 (2020), 2964–96.

[67] Thomas S Dee. "Teachers and the gender gaps in student achievement". In: *Journal of Human resources* 42.3 (2007), 528–554.

[68] Briggs Depew and Ozkan Eren. "Born on the wrong day? School entry age and juvenile crime". In: *Journal of Urban Economics* 96 (2016), 73–90.

[69] Eleanor Wiske Dillon and Jeffrey Andrew Smith. "Determinants of the match between student ability and college quality". In: *Journal of Labor Economics* 35.1 (2017), 45–66.

[70] Eleanor Wiske Dillon and Jeffrey Andrew Smith. "The consequences of academic match between students and colleges". In: *Journal of Human Resources* 55.3 (2020), 767–808.

[71] Lester E Dubins and David A Freedman. "Machiavelli and the Gale-Shapley algorithm". In: *The American Mathematical Monthly* 88.7 (1981), 485–494.

[72] Glenn Ellison and Parag A Pathak. "The Efficiency of Race-Neutral Alternatives to Race-Based Affirmative Action: Evidence from Chicago's Exam Schools". In: *American Economic Review* 111.3 (2021), 943–75.

[73] Ozkan Eren, Briggs Depew, and Stephen Barnes. "Test-based promotion policies, dropping out, and juvenile crime". In: *Journal of Public Economics* 153 (2017), 9–31.

[74] Ozkan Eren, Michael F Lovenheim, and Naci H Mocan. *The effect of grade retention on adult crime: Evidence from a test-based promotion policy*. Tech. rep. National Bureau of Economic Research, 2018.

[75] Fernanda Estevan et al. "The Top-Ten Way to Integrate High Schools". In: (2017).

[76] European Institute for Crime Prevention and Control, affiliated with the United Nations. *International Statistics on Crime and Justice*. HEUNI Publication Series 64. Helsinki: United Nations, 2010.

[77] David K Evans, Maryam Akmal, Pamela Jakiela, et al. "Gender gaps in education: The long view". In: *Center for Global Development* (2020).

[78] Christine L Exley and Judd B Kessler. "The gender gap in self-promotion". In: *The Quarterly Journal of Economics* (2022).

[79] Gabrielle Fack, Julien Grenet, and Yinghua He. "Beyond truth-telling: Preference estimation with centralized school choice and college admissions". In: *American Economic Review* 109.4 (2019), 1486–1529.

[80] Jeffrey Fagan and Edward Pabon. "Contributions of delinquency and substance use to school dropout among inner-city youths". In: *Youth and Society* 21.3 (1990), 306.

[81] Eduardo Fajnzylber, Bernardo Lara, and Tomás León. "Increased learning or GPA inflation? Evidence from GPA-based university admission in Chile". In: *Economics of Education Review* 72 (2019), 147–165.

[82] Jane Cooley Fruehwirth, Salvador Navarro, and Yuya Takahashi. "How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects". In: *Journal of Labor Economics* 34.4 (2016), 979–1021.

[83] Roland G Fryer Jr and Glenn C Loury. "Affirmative action and its mythology". In: *Journal of Economic Perspectives* 19.3 (2005), 147–162.

[84] David Gale and Lloyd S Shapley. "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1 (1962), 9–15.

[85] Varun Gauri. *School Choice In Chile: Two Decades of Educational Refor*. Pitt Latin American Studies. University of Pittsburgh Press, 1999.

[86] Claudia Goldin. "A grand gender convergence: Its last chapter". In: *American Economic Review* 104.4 (2014), 1091–1119.

[87] Claudia Goldin. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton University Press, 2021.

[88] Joshua Goodman et al. "The relationship between siblings' college choices: Evidence from one million SAT-taking families". In: *Economics of Education Review* 48 (2015), 75–85.

[89] Andrew Goodman-Bacon. "Difference-in-differences with variation in treatment timing". In: *Journal of Econometrics* 225.2 (2021), 254–277.

[90] N. Grau and D. Vergara. "An Observational Implementation of the Outcome Test with an Application to Ethnic Prejudice in Pretrial Detentions". In: *Working Paper* (2021).

[91] Nicolas Grau, Daniel Hojman, and Alejandra Mizala. "School closure and educational attainment: Evidence from a market-based system". In: *Economics of Education Review* 65 (2018), 1 –17. ISSN: 0272-7757. `https://doi.org/https://doi.org/10.1016/j.econedurev.2018.05.003`. `http://www.sciencedirect.com/science/article/pii/S0272775717303953`.

[92] Jay P Greene and Marcus A Winters. "The effects of exemptions to Florida's test-based promotion policy: Who is retained?: Who benefits academically?" In: *Economics of Education Review* 28.1 (2009), 135–142.

[93] Ingrid Haegele. "Talent Hoarding in Organizations". In: *Working Paper* (2022).

[94] Ingrid Haegele. "The Broken Rung: Gender and the Leadership Gap". In: *Working Paper* (2021).

[95] Guillaume Haeringer and Flip Klijn. "Constrained school choice". In: *Journal of Economic theory* 144.5 (2009), 1921–1947.

[96] Sukjin Han and Edward J. Vytlacil. "Identification in a generalization of bivariate probit models with dummy endogenous regressors". In: *Journal of Econometrics* 199.1 (2017), 63 –73. ISSN: 0304-4076. `https://doi.org/https://doi.org/10.1016/j.jeconom.2017.04.001`. `http://www.sciencedirect.com/science/article/pii/S0304407617300465`.

[97] Moritz Hardt, Eric Price, and Nati Srebro. "Equality of opportunity in supervised learning". In: *Advances in neural information processing systems* 29 (2016), 3315–3323.

[98] Justine Hastings, Thomas J Kane, and Douglas O Staiger. "Heterogeneous preferences and the efficacy of public school choice". In: *NBER Working Paper* 2145 (2009), 1–46.

[99] Justine S Hastings, Christopher A Neilson, and Seth D Zimmerman. *Are some degrees worth more than others? Evidence from college admission cutoffs in Chile*. Tech. rep. National Bureau of Economic Research, 2013.

[100] Zoltán Hermann and Marianna Kopasz. "Educational policies and the gender gap in test scores: a cross-country analysis". In: *Research Papers in Education* 36.4 (2021), 461–482.

[101] Paul Hirschfield. "Another way out: The impact of juvenile arrests on high school dropout". In: *Sociology of Education* 82.4 (2009), 368–393.

[102] Mark Hoekstra. "The effect of attending the flagship state university on earnings: A discontinuity-based approach". In: *The review of economics and statistics* 91.4 (2009), 717–724.

[103] C Thomas Holmes et al. “Grade level retention effects: A meta-analysis of research studies”. In: *Flunking grades: Research and policies on retention* 16 (1989), 33.

[104] Harry Holzer and David Neumark. “Assessing affirmative action”. In: *Journal of Economic literature* 38.3 (2000), 483–568.

[105] Michael Hout. “Social and economic returns to college education in the United States”. In: *Annual review of sociology* 38 (2012), 379–400.

[106] P. Hull. “What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making”. In: *Working Paper* (2021).

[107] Federico Huneeus et al. “Firm Sorting, College Major, and the Gender Earnings Gap”. In: *Working Paper* (2021).

[108] Joshua Hyman. “ACT for all: The effect of mandatory college entrance exams on postsecondary attainment and choice”. In: *Education Finance and Policy* 12.3 (2017), 281–311.

[109] Brian Jacob and Jesse Rothstein. “The measurement of student ability in modern assessment systems”. In: *Journal of Economic Perspectives* 30.3 (2016), 85–108.

[110] Brian A. Jacob. “Accountability, Incentives and Behavior: Evidence from School Reform in Chicago”. In: *Journal of Public Economics* 89 (2005), 761–796.

[111] Brian A. Jacob and Lars Lefgren. “The Effect of Grade Retention on High School Completion”. In: *American Economic Journal: Applied Economics* 1.3 (2009), 33–58.

[112] Shane R Jimerson. “Meta-analysis of grade retention research: Implications for practice in the 21st century”. In: *School psychology review* 30.3 (2001), 420.

[113] Adam Kapor et al. *Distributional effects of race-blind affirmative action*. Tech. rep. 2020.

[114] Elizabeth M King, Peter F Orazem, and Elizabeth M Paterno. “Promotion with and without learning: Effects on student enrollment and dropout behavior”. In: *The World Bank Economic Review* (2015), lhv049.

[115] Lars J Kirkeboen, Edwin Leuven, and Magne Mogstad. “Field of study, earnings, and self-selection”. In: *The Quarterly Journal of Economics* 131.3 (2016), 1057–1111.

[116] Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. "Inherent trade-offs in algorithmic fairness". In: *Proceedings of Innovations in Theoretical Computer Science*. 2017.

[117] Jon Kleinberg et al. "Algorithmic fairness". In: *Aea papers and proceedings*. Vol. 108. 2018, 22–27.

[118] J. Knowles, N. Persico, and P. Todd. "Racial bias in motor vehicle searches: Theory and evidence". In: *Journal of Political Economy* 109.1 (2001), 203–229.

[119] Martin Foureaux Koppensteiner. "Automatic grade promotion and student performance: Evidence from Brazil". In: *Journal of Development Economics* 107 (2014), 277 –290. ISSN: 0304-3878.

[120] Alan Krueger, Jesse Rothstein, and Sarah Turner. "Race, income, and college in 25 years: Evaluating Justice O'Connor's conjecture". In: *American Law and Economics Review* 8.2 (2006), 282–311.

[121] Jeanne Lafortune, Nicolás Figueroa, and Alejandro Saenz. *Do you like me enough? The impact of restricting preferences ranking in a university matching process*. Tech. rep. Working Paper, 2016.

[122] Rasmus Landersø, Helena Skyt Nielsen, and Marianne Simonsen. "School Starting Age and the Crime-age Profile". In: *The Economic Journal* (2016). DOI: 10.1111/ecoj.123251[]DOI: 10.1111/ecoj.12325. ISSN: 1468-0297.

[123] Máximo Langer and Ricardo Lillo. "Reforma a la justicia penal juvenil y adolescentes privados de libertad en Chile: Aportes empíricos para el debate". In: *Política criminal* 9 (2014), 713 –738.

[124] Tomás Larroucau and Ignacio Rios. "Do "short-list" students report truthfully? Strategic behavior in the Chilean college admissions problem". In: *Preprint, submitted September* 1.10.13140 (2018).

[125] Tomás Larroucau and Ignacio Rios. *Dynamic college admissions and the determinants of students' college retention*. 2020.

[126] Tomás Larroucau, Ignacio Ríos, and Alejandra Mizala. "Efecto de la incorporación del ranking de notas en el proceso de admisión a las universidades chilenas". In: *Pensamiento Educativo, Revista de Investigación Latinoamericana (PEL)* 52.1 (2015), 95–118.

[127] Chuhui Li, D.S. Poskitt, and Xueyan Zhao. "The bivariate probit model, maximum likelihood estimation, pseudo true parameters and partial identification". In: *Journal of Econometrics* 209.1 (2019), 94 –113. ISSN: 0304-4076. https://doi.org/https://doi.org/10.1016/j.jeconom.2018.07.009. http://www.sciencedirect.com/science/article/pii/S0304407618302562.

[128] Jaegeum Lim and Jonathan Meer. "Persistent effects of teacher–student gender matches". In: *Journal of Human Resources* 55.3 (2020), 809–835.

[129] Jaegeum Lim and Jonathan Meer. "The impact of teacher–student gender matches random assignment evidence from South Korea". In: *Journal of Human Resources* 52.4 (2017), 979–997.

[130] Lena Lindahl and Håkan Regnér. "College choice and subsequent earnings: Results using Swedish sibling data". In: *Scandinavian Journal of Economics* 107.3 (2005), 437–457.

[131] Lance Lochner. "Education, Work, And Crime: A Human Capital Approach". In: *International Economic Review* 45.3 (Aug. 2004), 811–843.

[132] Lance Lochner and Enrico Moretti. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports". In: *American Economic Review* 94.1 (2004), 155–189.

[133] Linda Datcher Loury and David Garman. "Affirmative action in higher education". In: *The American Economic Review* 83.2 (1993), 99–103.

[134] S. Machin, O. Marie, and S. Vujić. "The crime reducing effect of education". In: *The Economic Journal* 121 (2011), 463–484.

[135] Marco Manacorda. "The Cost of Grade Retention". In: *The Review of Economics and Statistics* 94.2 (2012), 596–606.

[136] Jessica P Marini et al. "Differential Validity and Prediction of the SAT®: Examining First-Year Grades and Retention to the Second Year." In: *College Board* (2019).

[137] Joan Martinez. "The Long-Term Effects of Teachers' Gender Bias". In: *arXiv preprint arXiv:2212.08220* (2022).

[138] Krista Mattern, Edgar Sanchez, and Edwin Ndum. "Why do achievement measures underpredict female academic performance?" In: *Educational Measurement: Issues and Practice* 36.1 (2017), 47–57.

[139] Krista D Mattern et al. "Differential Validity and Prediction of the SAT®. Research Report No. 2008-4." In: *College Board* (2008).

[140] Lorena Meckes and Rafael Carrasco. "Two decades of SIMCE: an Overview of the National Assessment System in Chile". In: *Assessment in Education: Principles, Policy and Practice* 17.2 (2010), 233–248. https://doi.org/10.1080/09695941003696214. eprint: http://www.tandfonline.com/doi/pdf/10.1080/09695941003696214. http://www.tandfonline.com/doi/abs/10.1080/09695941003696214.

[141] Ursula Mello. "Affirmative Action and the Choice of Schools". In: (2021).

[142] Ursula Mello. "Centralized Admissions, Affirmative Action and Access of Low-income Students to Higher Education". In: *American Economic Journal: Economic Policy* (2021).

[143] Ursula Mello. "Centralized Admissions, Affirmative Action, and Access of Low-Income Students to Higher Education". In: *American Economic Journal: Economic Policy* 14.3 (2022), 166–97.

[144] Jack Mountjoy. "Community colleges and upward mobility". In: *Working Paper* (2021).

[145] Jack Mountjoy and Brent R Hickman. *The Returns to College (s): Relative Value-Added and Match Effects in Higher Education.* Tech. rep. National Bureau of Economic Research, 2021.

[146] Sendhil Mullainathan and Ziad Obermeyer. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care". In: *Working Paper* (2021).

[147] Karthik Muralidharan and Ketki Sheth. "Bridging education gender gaps in developing countries: The role of female teachers". In: *Journal of Human Resources* 51.2 (2016), 269–297.

[148] Muriel Niederle and Lise Vesterlund. "Do women shy away from competition? Do men compete too much?" In: *The Quarterly Journal of Economics* 122.3 (2007), 1067–1101.

[149] Muriel Niederle and Lise Vesterlund. "Explaining the gender gap in math test scores: The role of competition". In: *Journal of Economic Perspectives* 24.2 (2010), 129–44.

[150] Sebastián Otero, Nano Barahona, and Cauê Dobbin. *Affirmative action in centralized college admission systems: Evidence from Brazil.* Tech. rep. Working paper, 2021.

[151] Suh-Ruu Ou and Arthur J Reynolds. "Grade retention, postsecondary education, and public aid receipt". In: *Educational Evaluation and Policy Analysis* 32.1 (2010), 118–139.

[152] Valentina Paredes. “A teacher like me or a student like me? Role model versus teacher bias effect”. In: *Economics of Education Review* 39 (2014), 38–49.

[153] Parag A Pathak and Tayfun Sönmez. “School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation”. In: *American Economic Review* 103.1 (2013), 80–106.

[154] Devin G Pope and Justin R Sydnor. “Geographic variation in the gender differences in test scores”. In: *Journal of Economic Perspectives* 24.2 (2010), 95–108.

[155] Ashesh Rambachan. “Identifying prediction mistakes in observational data”. In: *Working Paper* (2021).

[156] Ashesh Rambachan et al. “An economic approach to regulating algorithms”. In: *Working Paper* (2021).

[157] Amy L Reschly and Sandra L Christenson. “Grade retention: Historical perspectives and new research”. In: *Journal of school psychology* 51.3 (2013), 319–322.

[158] Ignacio Rios et al. “Improving the Chilean college admissions system”. In: *Operations Research* 69.4 (2021), 1186–1205.

[159] Melissa Roderick. “Grade retention and school dropout: Investigating the association”. In: *American Educational Research Journal* 31.4 (1994), 729–759.

[160] Alvin E Roth. “The economics of matching: Stability and incentives”. In: *Mathematics of operations research* 7.4 (1982), 617–628.

[161] Jesse Rothstein. “Qualitative Information in Undergraduate Admissions: A Pilot Study of Letters of Recommendation”. In: *Working Paper* (2021).

[162] Jesse Rothstein and Albert H Yoon. *Affirmative action in law school admissions: What do racial preferences do?* Tech. rep. National Bureau of Economic Research, 2008.

[163] Jesse M Rothstein. “College performance predictions and the SAT”. In: *Journal of Econometrics* 121.1-2 (2004), 297–317.

[164] Richard Sander and Stuart Taylor. *Mismatch: How Affirmative Action Hurts Students It’s Intended to Help, and Why Universities Won’t Admit It*. Basic Books, 2012.

[165] H. Sarsons. “Interpreting Signals in the Labor Market: Evidence from Medical Referrals”. In: *Working Paper* (2019).

[166] Perihan O Saygin and Ann Atwater. "Gender differences in leaving questions blank on high-stakes standardized tests". In: *Economics of Education Review* 84 (2021), 102162.

[167] Camelia Simoiu, Sam Corbett-Davies, and Sharad Goel. "The problem of infra-marginality in outcome tests for discrimination". In: *The Annals of Applied Statistics* 11.3 (2017), 1193–1216.

[168] Carolyn M Sloane, Erik G Hurst, and Dan A Black. "College Majors, Occupations, and the Gender Wage Gap". In: *Journal of Economic Perspectives* 35.4 (2021), 223–48.

[169] Alex Solis. *The Effects of Grade Retention on Human Capital Accumulation.* Working Paper Series 2017:15. Uppsala University, Department of Economics, Mar. 2017. `https://ideas.repec.org/p/hhs/uunewp/2017_015.html`.

[170] Thomas Sowell. *Black Education: Myths and Tragedies.* David McKay, 1972.

[171] Nicholas Turner. "Income Segregation and Intergenerational Mobility Across Colleges in the United States". In: (2020).

[172] Sarah E Turner and William G Bowen. "Choice of major: The changing (unchanging) gender gap". In: *ILR Review* 52.2 (1999), 289–313.

[173] Paul A Westrick et al. "Validity of the SAT® for predicting first-year grades and retention to the second year". In: *College Board Research Paper* (2019).

[174] Danny Yagan. "Supply vs. demand under an affirmative action ban: Estimates from UC law schools". In: *Journal of Public Economics* 137 (2016), 38–50.

[175] Seth D Zimmerman. "Elite colleges and upward mobility to top jobs and top incomes". In: *American Economic Review* 109.1 (2019), 1–47.

[176] Seth D Zimmerman. "The returns to college admission for academically marginal students". In: *Journal of Labor Economics* 32.4 (2014), 711–754.

[177] Rebecca Zwick and Jennifer Greif Green. "New perspectives on the correlation of SAT scores, high school grades, and socioeconomic factors". In: *Journal of Educational Measurement* 44.1 (2007), 23–45.

Appendix A

Appendix of The impact of grade retention on juvenile crime

A.1 Additional figures and tables

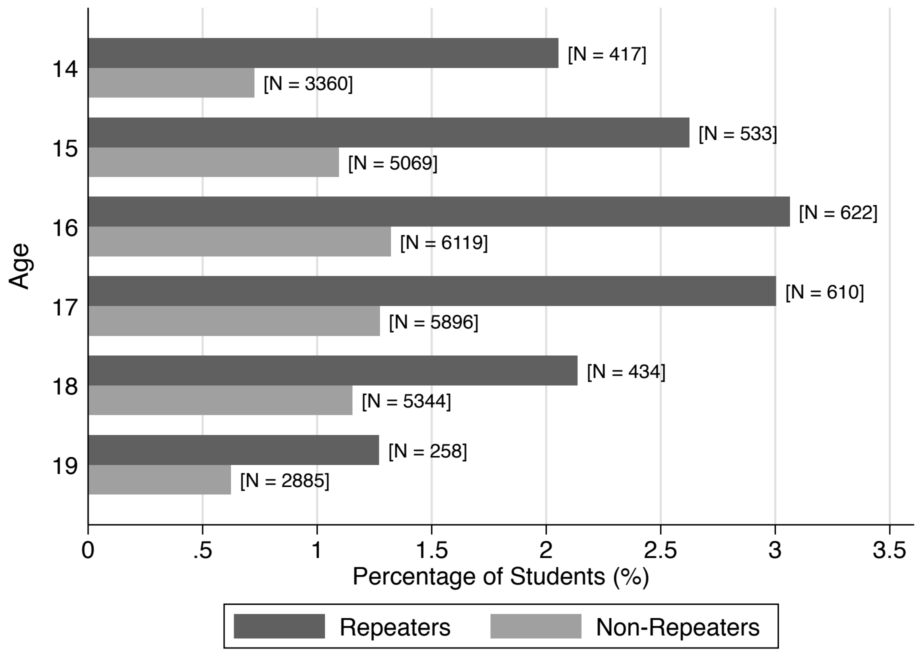
Descriptive statistics for juvenile crime

Table A.1: Juvenile crime distribution

Crime Category	Freq.	Percentage	Pretrial Detention (%)
Theft	1057	24.34	0.76
Non-violent Robbery	662	15.25	11.03
Other Crimes Agains Property	567	13.06	4.23
Robbery	522	12.02	29.89
Injuries	394	9.07	1.78
Crimes Against Sexual Freedom and Privacy	343	7.90	3.50
Other Crimes	274	6.31	1.46
Offenses	169	3.89	0.59
Crimes Against Drug Laws	123	2.83	7.32
Crimes Against Special Laws	67	1.54	29.85
Traffic Law Crimes	52	1.20	0.00
Sex Crimes	45	1.04	8.89
Crimes Against Public Faith	18	0.41	5.56
Homicides	18	0.41	44.44
Intellectual and Industrial Property Crimes	13	0.30	7.69
Financial and Tax Crimes	9	0.21	0.00
Crimes Against Millitary Laws	5	0.12	20.00
Negligent Offense	3	0.07	0.00
Facts of Criminal Relevance	1	0.02	0.00

Notes: This plot shows the distribution of crimes for students who attended 2nd or 3rd grade in the year 2007. Severe crimes are the ones in which case the pretrial detention rate is greater than 3%.

Figure A.1: % of students who were criminally prosecuted by age



Notes: This table shows, among repeaters and non-repeaters in 2nd and 3rd grades (2007), the fraction of students prosecuted for the first time at different ages.

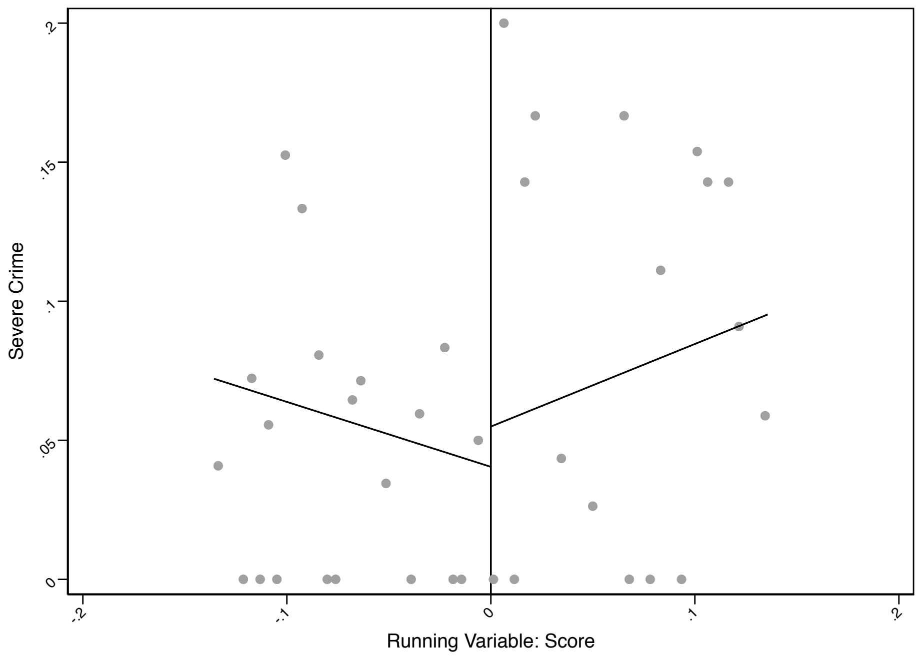
More results

Table A.2: Effects of Grade Retention: reduced form

	All Crimes	Severe Crimes	Dropout	Future Grade Retention	GPA From 4th to 8th Grade
	(1)	(2)	(3)	(4)	(5)
RD Estimator	.071**	.056*	.225***	.293***	-.198***
	(.027)	(.025)	(.058)	(.052)	(.04)
Mean Variable	.132	.106	.508	.645	5.163
Std. Dev. Variable	.027	.025	.058	.052	.04
Robust Inference					
p-value	0.02	0.05	0.00	0.00	0.00
C.I.	[.01 .132]	[0 .113]	[.1 .35]	[.179 .408]	[-.284 -.112]
Effective Obs.					
Left	1,845	1,618	829	1,062	898
Right	610	571	415	470	407
Optimal Bandwidth[a]	.193	.177	.104	.128	.12

Notes: This table presents the results for the impact of grade retention on the 5 listed outcomes, based on the methods for estimation and inference for sharp RD designs. Note that in this case, the sign is reversed because now we are accounting for the effect of crossing the score cutoff (which decreases the probability of grade retention). Therefore the interpretation here is: by getting an average score above the cutoff, and therefore decreasing their probability of grade retention, the students are, on average, 7 pp more likely to commit a crime, 22 pp more likely to drop out and 29 pp more likely of being retained a grade in the future. The robust inference considers the bias term coming from the approximation error that does not vanish from the asymptotic distribution of the RD estimator. Standard errors are in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

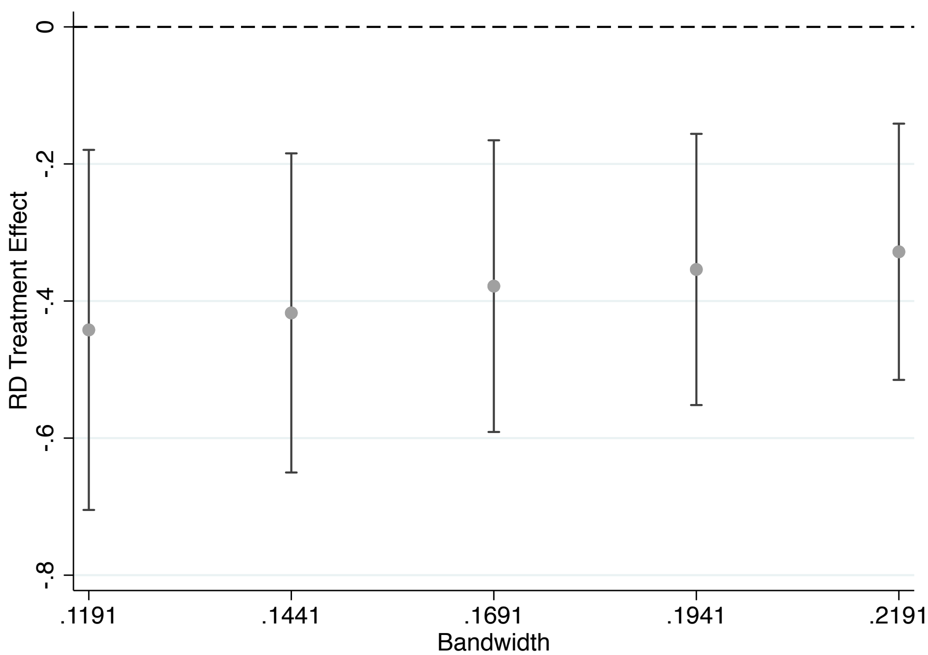
Figure A.2: Graphic results for severe crime



Notes: This figure shows the values of the outcome and an estimation of the regression functions via local linear regressions around the threshold for grade retention for the case of severe crimes.

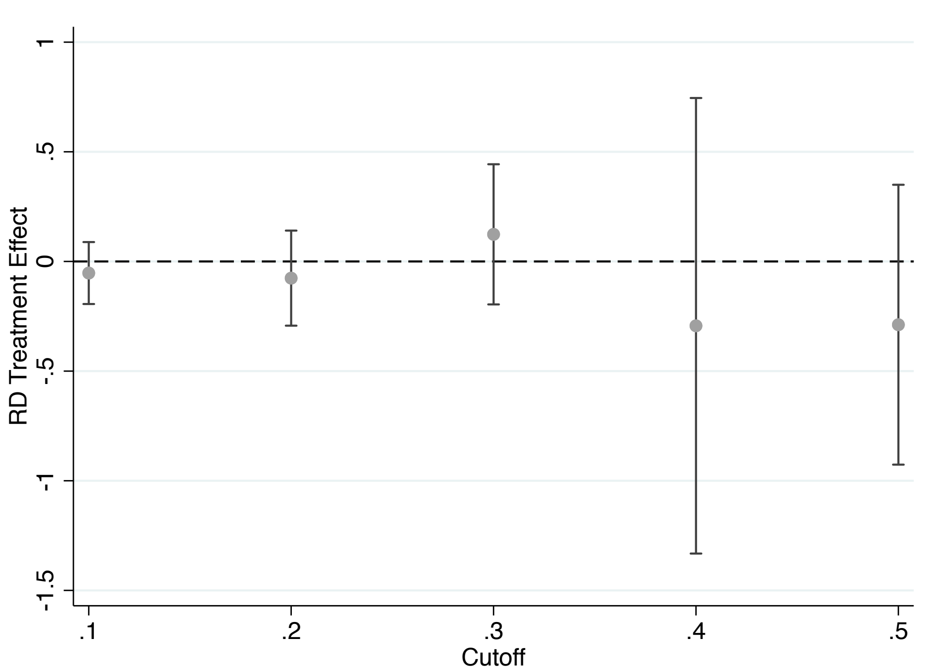
Robustness Analysis

Figure A.3: Sensitivity to Bandwidth: Dropout



Notes: This figure shows the fuzzy RD estimations for the impact of grade retention on juvenile crime, for different values of the bandwidth (the third estimate is the one with the optimal bandwidth). The point estimates are the dots and the confidence intervals at 95% are the brackets.

Figure A.4: Placebo Tests: Dropout



Notes: This figure shows the sharp RD estimations for the impact of being below the cutoff on juvenile crime, for different values of the cutoff. The point estimates are the dots and the confidence intervals at 95% are the brackets.

Model's estimated parameters

Restricted sample

Table A.3: GPA estimated parameters (Restricted Sample)

	1st period GPA		2nd period GPA	
	α	S.E.	α	S.E.
Male	-0.007	0.004	-0.123	0.020
3rd Grade in 2007	-0.017	0.005	-0.086	0.026
Missing mother educ	0.029	0.006	-0.065	0.028
Mother educ. >14 years	0.011	0.009	0.036	0.044
Sch. Average: Father educ.	-0.001	0.003	0.009	0.013
Sch. Average: Mother educ.	0.005	0.003	0.002	0.014
Sch. Average: Math SIMCE	-0.0003	0.0001	0.003	0.001
Sch. Average: Spanish SIMCE	0.0001	0.0001	-0.001	0.001
Public School	0.005	0.005	0.088	0.023
Repeated in period one (type I)	.	.	0.132	0.027
Repeated in period one (type II)	.	.	-1.146	0.179
Constant Type I	4.883	0.029	4.076	0.142
Constant Type II	4.892	0.036	3.072	0.195
Log(Standard Error)	-2.526	0.023	-0.891	0.016

Notes: This table presents the point estimates and standard errors for the parameters of equations (1.1) and (1.2), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Table A.4: Grade retention, crime, and types estimated parameters (Restricted Sample)

	Grade Retention			Crime	
	γ	S.E.		β	S.E.
First Period:			Repeated in 1st period	-0.141	0.203
$1(GPA < 4.95)$	1.545	0.149	Repeated in 2nd period	0.318	0.205
$GPA - 4.95$	-3.183	0.893	Repeated both periods	-0.058	0.230
$(GPA - 4.95)^2$	24.475	5.926	Male	0.504	0.106
3rd Grade in 2007	-0.316	0.085	3rd Grade in 2007	0.132	0.121
Constant Type I	-0.409	0.117	Missing mother educ	0.017	0.133
Constant Type II	-1.400	0.287	Mother educ. >14 years	0.344	0.242
			First Period variables:		
Second Period:			GPA	-0.442	0.809
Repeated in 1st period	-0.673	0.084	Sch. Average: Father educ.	0.020	0.099
GPA	-1.410	1.267	Sch. Average: Mother educ.	0.079	0.096
GPA^2	0.130	0.137	Sch. Average: Math SIMCE	-0.010	0.006
$1(GPA < 4.45)$	0.755	0.129	Sch. Average: Spanish SIMCE	0.003	0.005
$1(GPA < 4.95)$	1.124	0.160	Public School	0.240	0.151
3rd Grade in 2007	-0.102	0.071	**Second Period variables:**		
Constant Type I	2.834	2.974	GPA	-0.631	0.140
Constant Type II	0.422	2.446	Sch. Average: Father educ.	-0.004	0.101
			Sch. Average: Mother educ.	-0.197	0.099
Type distribution:			Sch. Average: Math SIMCE	0.000	0.006
Type I parameter	4.209	0.263	Sch. Average: Spanish SIMCE	0.002	0.005
Type I probability	0.99	.	Public School	-0.169	0.160
Type II probability	0.01	.	Constant Type I	5.818	4.114
			Constant Type II	5.252	4.125

Notes: This table presents the point estimates and standard errors for the parameters of equations (1.3), (1.4), and (1.5), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Full sample

Table A.5: GPA estimated parameters (Full Sample)

	1st period GPA		2nd period GPA	
	α	S.E.	α	S.E.
Male	-0.083	0.007	-0.116	0.009
3rd Grade in 2007	-0.009	0.008	-0.087	0.011
Missing mother educ	0.033	0.009	-0.038	0.011
Mother educ. >14 years	0.002	0.018	0.077	0.024
Sch. Average: Father educ.	0.012	0.005	-0.012	0.006
Sch. Average: Mother educ.	0.011	0.005	0.026	0.006
Sch. Average: Math SIMCE	0.001	0.0002	0.002	0.0004
Sch. Average: Spanish SIMCE	0.001	0.0002	-0.0001	0.0004
Public School	-0.015	0.007	0.029	0.009
Repeated at period one (type I)	.	.	0.170	0.157
Repeated at period one (type II)	.	.	0.080	0.019
	0.308			
Constant Type I	3.349	0.047	3.680	0.047
Constant Type II	3.955	0.046	4.019	0.061
Log(Standard Error)	-1.209	0.011	-0.775	0.005

Notes: This table presents the point estimates and standard errors for the parameters of equations (1.1) and (1.2), using the full sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Table A.6: Grade retention, crime, and types estimated parameters (Full Sample)

	Grade Retention			Crime	
	γ	S.E.		β	S.E.
First Period:			Repeated in 1st period	0.005	0.111
$1(GPA < 4.95)$	1.853	0.085	Repeated in 2nd period	0.269	0.127
$GPA - 4.95$	-1.321	0.152	Repeated both periods	-0.021	0.131
$(GPA - 4.95)^2$	-0.426	0.125	Male	0.475	0.036
3rd Grade in 2007	-0.289	0.046	3rd Grade in 2007	0.053	0.037
Constant Type I	-0.125	0.274	Missing mother educ	0.015	0.039
Constant Type II	-0.323	0.061	Mother educ. >14 years	0.040	0.088
			First Period variables:		
Second Period:			GPA	-0.059	0.097
Repeated in 1st period	-0.347	0.050	Sch. Average: Father educ.	0.021	0.028
GPA	0.034	0.280	Sch. Average: Mother educ.	0.010	0.028
GPA^2	0.016	0.034	Sch. Average: Math SIMCE	-0.002	0.002
$1(GPA < 4.45)$	1.083	0.042	Sch. Average: Spanish SIMCE	0.000	0.002
$1(GPA < 4.95)$	1.177	0.064	Public School	0.040	0.043
3rd Grade in 2007	-0.069	0.026	**Second Period variables:**		
Constant Type I	-1.313	0.569	GPA	-0.386	0.036
Constant Type II	-1.727	0.572	Sch. Average: Father educ.	-0.030	0.027
			Sch. Average: Mother educ.	-0.040	0.028
Type distribution:			Sch. Average: Math SIMCE	-0.001	0.002
Type I parameter	3.296	0.141	Sch. Average: Spanish SIMCE	-0.001	0.002
Type II parameter	4.882	0.139	Public School	0.080	0.043
Type I probability	0.17	.	Constant Type I	2.077	0.465
Type II probability	0.82	.	Constant Type II	1.935	0.553

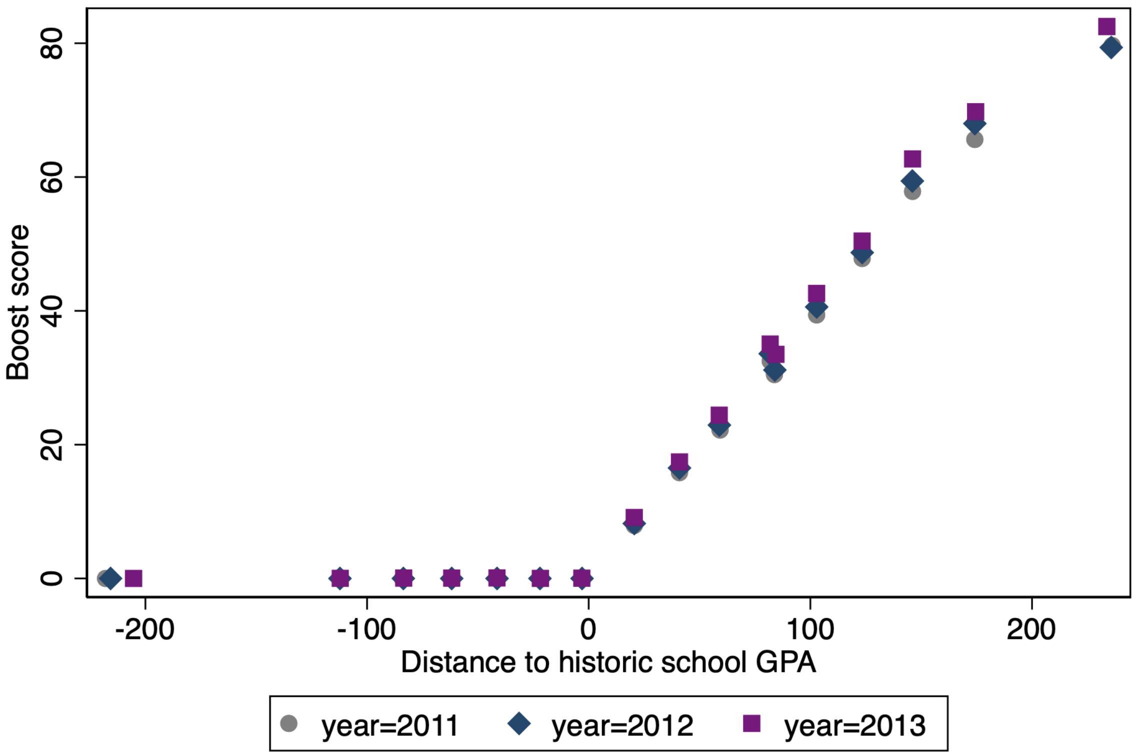
Note: This table presents the point estimates and standard errors for the parameters of equations (1.3), (1.4), and (1.5), using the restricted sample. The standard errors are calculated using the approximation of the Hessian given by the mean of the outer product of the scores.

Appendix B

Appendix of The equity and efficiency effects of a relative GPA reward in college admissions

B.1 Additional figures and tables

Figure B.1: Boost score by relative position of student in high school



Notes: boost score for cohort 2011, 2012, and 2013. For 2013 GPA$^+$ (and the inferred boost) was provided on the application data. For 2011 and 2012 boost was calculated according to the GPA$^+$ formula using education records of the universe of high school students who graduated between 2008 and 2012.

Table B.1: Enrollment rates at admission program by groups, before and after the reform

Total	**Unaffected**	**Pulled-Up**	**Pushed-Down**
Enrollment Pre-Reform (2012)	0.80	0.83	0.91
Enrollment Reform (2013)	0.79	0.87	0.85
Difference	-0.01	0.04	-0.06

Program	**Unaffected**	**Pulled-Up**	**Pushed-Down**
Enrollment Pre-Reform (2012)	0.60	0.53	0.78
Enrollment Reform (2013)	0.62	0.75	0.66
Difference	0.02	0.22	-0.12

Non-selective	**Unaffected**	**Pulled-Up**	**Pushed-Down**
Enrollment Pre-Reform (2012)	0.13	0.12	0.08
Enrollment Reform (2013)	0.10	0.05	0.08
Difference	-0.03	-0.07	0.00

Notes: averages for a variable that indicates if the student chooses to enroll in the admission assignment. The difference by group, between after and before the reform is shown in the 3rd row.

Table B.2: Effect on graduation from admission program conditional on enrollment

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	-0.012	-0.009	-0.010	-0.000
	(0.012)	(0.013)	(0.014)	(0.014)
P-Down x after	-0.005	-0.018	-0.034***	0.001
	(0.010)	(0.012)	(0.012)	(0.012)
Obs.	144,540	144,540	144,540	144,540
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.302	0.126	0.020	0.989

Notes: Diff-in-diff results for the sample of students that enroll in 1st year. Columns 1-3 show estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time.

Table B.3: Graduation averages from any program by groups, before and after the reform

	Unaffected	**Pulled-Up**	**Pushed-Down**
Grad by 6yr Pre-Reform (2012)	0.22	0.24	0.21
Grad by 6yr Reform (2013)	0.21	0.22	0.19
Difference	-0.01	-0.02	-0.02

	Unaffected	**Pulled-Up**	**Pushed-Down**
Grad by 7yr Pre-Reform (2012)	0.36	0.40	0.35
Grad by 7yr Reform (2013)	0.34	0.37	0.34
Difference	-0.02	-0.03	-0.01

	Unaffected	**Pulled-Up**	**Pushed-Down**
Grad by 8yr Pre-Reform (2012)	0.46	0.51	0.47
Grad by 8yr Reform (2013)	0.42	0.45	0.42
Difference	-0.04	-0.06	-0.05

Notes: averages for a variable that indicates if the student graduates from some program (selective or non-selective). The difference by group, between after and before the reform is shown in the 3rd row.

Heterogeneity Analysis

Table B.4: Heterogeneity: Effects on enrollment by gender, family income, and boost

	(1) Enrollment	(2) Enrollment	(3) Enrollment
P-Up x after	0.193***	0.192***	0.080***
	(0.017)	(0.014)	(0.029)
P-Down x after	-0.148***	-0.145***	-0.148***
	(0.013)	(0.011)	(0.012)
P-Up x after x Characteristic	0.009	0.020	0.155***
	(0.022)	(0.021)	(0.031)
P-Down x after x Characteristic	0.027	0.033	0.012
	(0.020)	(0.022)	(0.020)
After x Characteristic	-0.005	-0.002	-0.023***
	(0.004)	(0.004)	(0.004)
Obs.	234,544	234,544	234,544
Controls	✓	✓	✓
Characteristic	Female	Low Income	Boost
PU Fraction	61%	45%	85%
PD Fraction	41%	30%	32%

Notes: main diff-in-diff specification for enrollment fully interacted with (i) female indicator, (ii) low-income indicator, and (iii) boost indicator. Robust standard errors are in parentheses.

Table B.5: Heterogeneity: Effects on graduation from the same program by gender, family income, and boost

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr
P-Up x after	0.042***	0.082***	0.025
	(0.014)	(0.013)	(0.022)
P-Down x after	-0.051***	-0.058***	-0.055***
	(0.012)	(0.012)	(0.011)
P-Up x after x Characteristic	0.047**	-0.025	0.065***
	(0.019)	(0.020)	(0.025)
P-Down x after x Characteristic	-0.020	-0.013	-0.049**
	(0.020)	(0.021)	(0.021)
After x Characteristic	0.010***	0.006*	-0.015***
	(0.004)	(0.004)	(0.004)
Obs.	234,544	234,544	234,544
Controls	✓	✓	✓
Characteristic	Female	Low Income	Boost
PU Fraction	61%	45%	85%
PD Fraction	41%	30%	32%

Notes: main diff-in-diff specification for graduation from the assigned program fully interacted with (i) female indicator, (ii) low-income indicator, and (iii) boost indicator. Robust standard errors are in parentheses.

Table B.6: Heterogeneity: Effects on graduation from any program by gender, family income, and boost

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr
P-Up x after	-0.023	-0.008	0.006
	(0.017)	(0.015)	(0.028)
P-Down x after	-0.004	-0.002	-0.008
	(0.014)	(0.013)	(0.013)
P-Up x after x Characteristic	0.003	-0.030	-0.018
	(0.023)	(0.023)	(0.031)
P-Down x after x Characteristic	-0.008	-0.026	-0.042*
	(0.022)	(0.024)	(0.023)
After x Characteristic	0.003	0.011***	-0.015***
	(0.004)	(0.004)	(0.004)
Obs.	234,544	234,544	234,544
Controls	✓	✓	✓
Characteristic	Female	Low Income	Boost
PU Fraction	61%	45%	85%
PD Fraction	41%	30%	32%

Notes: main diff-in-diff specification for graduation from any program fully interacted with (i) female indicator, (ii) low-income indicator, and (iii) boost indicator. Robust standard errors are in parentheses.

Table B.7: Heterogeneity: Effects on graduation from a selective program by gender, family income, and boost

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr
P-Up x after	-0.015	0.006	0.038
	(0.017)	(0.015)	(0.026)
P-Down x after	-0.006	-0.007	-0.022*
	(0.014)	(0.013)	(0.013)
P-Up x after x Characteristic	0.026	-0.010	-0.025
	(0.022)	(0.022)	(0.029)
P-Down x after x Characteristic	-0.022	-0.035	-0.031
	(0.022)	(0.023)	(0.023)
After x Characteristic	0.018***	0.021***	-0.021***
	(0.004)	(0.004)	(0.004)
Obs.	234,544	234,544	234,544
Controls	✓	✓	✓
Characteristic	Female	Low Income	Boost
PU Fraction	61%	45%	85%
PD Fraction	41%	30%	32%

Notes: main diff-in-diff specification for graduation from a selective program fully interacted with (i) female indicator, (ii) low-income indicator, and (iii) boost indicator. Robust standard errors are in parentheses.

Table B.8: Heterogeneity: Effects on graduation or enrollment after 8 years from a selective program by gender, family income, and boost

	(1) Grad or enroll by 8 yr	(2) Grad or enroll by 8 yr	(3) Grad or enroll by 8 yr
P-Up x after	-0.015	-0.001	-0.030
	(0.017)	(0.013)	(0.029)
P-Down x after	0.028**	0.014	0.013
	(0.013)	(0.012)	(0.013)
P-Up x after x Characteristic	0.005	-0.021	0.037
	(0.021)	(0.021)	(0.031)
P-Down x after x Characteristic	-0.014	0.020	-0.013
	(0.020)	(0.022)	(0.020)
After x Characteristic	0.010**	0.013***	-0.007*
	(0.004)	(0.004)	(0.004)
Obs.	234,544	234,544	234,544
Controls	✓	✓	✓
Characteristic	Female	Low Income	Boost
PU Fraction	61%	45%	85%
PD Fraction	41%	30%	32%

Notes: main diff-in-diff specification for graduation or enrollment after 8 years from a selective program fully interacted with (i) female indicator, (ii) low-income indicator, and (iii) boost indicator. Robust standard errors are in parentheses.

Main results from Section 2.9, boost sensitivity

Table B.9: Diff-in-diff estimates for enrollment

	(1) Enrollment	(2) Enrollment	(3) Non-Select	(4) Non-Select
P-Up x after	0.197***	0.219***	-0.047***	-0.056***
	(0.011)	(0.011)	(0.007)	(0.007)
P-Down x after	-0.136***	-0.167***	0.023***	0.039***
	(0.010)	(0.010)	(0.006)	(0.006)
Obs.	233,789	233,789	233,789	233,789
Controls		✓		✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.000	0.000	0.014	0.064

Notes: columns 1 and 2 have estimates when the outcome is enrollment at the admission program. Columns 3 and 4 have estimates for an indicator if the student enrolls in a non-selective program. Columns 2 and 4 control for standardized test scores, GPA, family income, region, type of high school, and gender. Robust standard errors are in parentheses.

Table B.10: Diff-in-diff estimates for graduation from GPA$^+$ program on sample without boost > 150

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.041***	0.070***	0.083***	0.041***
	(0.008)	(0.009)	(0.010)	(0.010)
P-Down x after	-0.033***	-0.060***	-0.082***	-0.038***
	(0.007)	(0.009)	(0.009)	(0.010)
Obs.	233,789	233,789	233,789	233,789
Controls	✓	✓	✓	✓
Test	0	0	0	0
p-value	0.483	0.418	0.968	0.852

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table B.11: Diff-in-diff estimates for any graduation on the sample without boost > 150

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8 yr
P-Up x after	-0.005	0.005	-0.007	0.007
	(0.010)	(0.011)	(0.011)	(0.010)
P-Down x after	-0.012	-0.014	-0.032***	-0.006
	(0.009)	(0.010)	(0.011)	(0.010)
Obs.	233,789	233,789	233,789	233,789
Controls	✓	✓	✓	✓

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Main results from Section 2.9, long programs

Table B.12: Diff-in-diff estimates for enrollment on the sample without long programs

	(1) Enrollment	(2) Enrollment	(3) Non-Select	(4) Non-Select
P-Up x after	0.210***	0.238***	-0.067***	-0.078***
	(0.014)	(0.013)	(0.010)	(0.009)
P-Down x after	-0.141***	-0.177***	0.025***	0.043***
	(0.013)	(0.013)	(0.009)	(0.009)
Obs.	178,760	178,760	178,760	178,760
Controls		✓		✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.000	0.001	0.002	0.009

Robust standard errors in parentheses

Notes: results for the sample without students in programs with 6 or 7 expected years. Columns 1 and 2 have estimates when the outcome is enrollment at the admission program. Columns 3 and 4 have estimates for an indicator if the student enrolls in a non-selective program. Columns 2 and 4 control for standardized test scores, GPA, family income, region, type of high school, and gender. Robust standard errors are in parentheses.

Table B.13: Diff-in-diff estimates for graduation from GPA$^+$ program on sample without long programs

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.065***	0.084***	0.098***	0.061***
	(0.011)	(0.012)	(0.012)	(0.013)
P-Down x after	-0.051***	-0.077***	-0.087***	-0.061***
	(0.011)	(0.012)	(0.012)	(0.013)
Obs.	178,760	178,760	178,760	178,760
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.367	0.667	0.552	0.981

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table B.14: Diff-in-diff estimates for any graduation on the sample without long programs

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8 yr
P-Up x after	0.008	0.007	0.008	0.009
	(0.013)	(0.014)	(0.014)	(0.013)
P-Down x after	-0.030**	-0.034**	-0.040***	-0.008
	(0.012)	(0.013)	(0.014)	(0.013)
Obs.	178,760	178,760	178,760	178,760
Controls	✓	✓	✓	✓

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Main results from Section 2.9, extra long programs

Table B.15: Diff-in-diff estimates for enrollment on the sample without extra long programs

	(1) Enrollment	(2) Enrollment	(3) Non-Select	(4) Non-Select
P-Up x after	0.195***	0.216***	-0.052***	-0.060***
	(0.011)	(0.011)	(0.007)	(0.007)
P-Down x after	-0.136***	-0.167***	0.023***	0.038***
	(0.010)	(0.010)	(0.007)	(0.006)
Obs.	228,741	228,741	228,741	228,741
Controls		✓		✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.000	0.001	0.004	0.029

Notes: results for the sample without students in programs with 6 or 7 expected years. Columns 1 and 2 show estimates when the outcome is enrollment in the admission program. Columns 3 and 4 have estimates for an indicator if the student enrolls in a non-selective program. Columns 2 and 4 control for standardized test scores, GPA, family income, region, type of high school, and gender. Robust standard errors are in parentheses.

Table B.16: Diff-in-diff estimates for graduation from GPA^{+} program on sample without extra long programs

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.047***	0.066***	0.080***	0.042***
	(0.008)	(0.009)	(0.010)	(0.010)
P-Down x after	-0.034***	-0.055***	-0.072***	-0.035***
	(0.008)	(0.009)	(0.010)	(0.010)
Obs.	228,741	228,741	228,741	228,741
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.272	0.411	0.547	0.623

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table B.17: Diff-in-diff estimates for any graduation on the sample without extra long programs

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8 yr
P-Up x after	0.002	0.001	-0.003	0.005
	(0.010)	(0.011)	(0.011)	(0.010)
P-Down x after	-0.014	-0.011	-0.023**	-0.005
	(0.009)	(0.010)	(0.011)	(0.010)
Obs.	228,741	228,741	228,741	228,741
Controls	✓	✓	✓	✓

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Main results with clustered standard errors at school-year level

Table B.18: Diff-in-diff estimates for enrollment at admission program with school-year cluster standard errors

	(1) Enrollment	(2) Enrollment	(3) Non-Select	(4) Non-Select
P-Up x after	0.199***	0.219***	-0.049***	-0.057***
	(0.012)	(0.011)	(0.007)	(0.007)
P-Down x after	-0.136***	-0.167***	0.023***	0.039***
	(0.011)	(0.011)	(0.007)	(0.007)
Obs.	234,544	234,544	234,544	234,544
Controls		✓		✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.000	0.001	0.014	0.063

Notes: Columns 1 and 2 show estimates when the outcome is enrollment at the admission program. Columns 3 and 4 have estimates for an indicator if the student enrolls in a non-selective program. Columns 2 and 4 control for standardized test scores, GPA, family income, region, type of high school, and gender.

Table B.19: Diff-in-diff estimates for graduation at admission program with school-year cluster standard errors

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad on time
P-Up x after	0.042***	0.072***	0.084***	0.043***
	(0.008)	(0.009)	(0.010)	(0.011)
P-Down x after	-0.033***	-0.060***	-0.082***	-0.039***
	(0.008)	(0.009)	(0.010)	(0.010)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓
Test	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$	$\beta_3 = -\beta_4$
p-value	0.412	0.376	0.918	0.751

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table B.20: Diff-in-diff estimates for college graduation (any program) with school-year cluster standard errors

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad or enroll by 8 yr
P-Up x after	-0.003	0.005	-0.008	0.006
	(0.010)	(0.011)	(0.011)	(0.010)
P-Down x after	-0.012	-0.015	-0.032***	-0.006
	(0.009)	(0.010)	(0.011)	(0.010)
Obs.	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Table B.21: Diff-in-diff estimates for graduation with school-year cluster standard errors

	(1) Grad by 6yr	(2) Grad by 7yr	(3) Grad by 8yr	(4) Grad by 6yr	(5) Grad by 7yr	(6) Grad by 8yr
P-Up x after	0.019**	0.030***	0.019*	-0.022***	-0.025***	-0.026***
	(0.009)	(0.011)	(0.011)	(0.004)	(0.005)	(0.005)
P-Down x after	-0.022***	-0.028***	-0.046***	0.009**	0.013***	0.014***
	(0.008)	(0.010)	(0.010)	(0.004)	(0.004)	(0.005)
Obs.	234,544	234,544	234,544	234,544	234,544	234,544
Controls	✓	✓	✓	✓		✓
	Selective	Selective	Selective	Non-Selective	Non-Selective	Non-Selective

Notes: Columns 1-3 show diff-in-diff estimates for an indicator if the student graduate from the admission program by 6, 7, and 8 years after application. Column 4 shows the results for the outcome of graduation on time. Robust standard errors are in parentheses.

Regression Discontinuity Results

Table B.22: RD estimates on enrollment for crossing the threshold for 1st preference

	(1) Enroll	(2) Enroll	(3) Enroll	(4) Enroll	(5) Enroll
RD estimator	0.170***	0.185***	0.184***	0.178***	0.136***
	(0.009)	(0.008)	(0.007)	(0.005)	(0.012)
Obs.	90,205	90,205	90,205	90,205	15,623
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2013 sample of students at programs with excess demand. The outcome variable indicates if the students enrolled in the admission program. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.23: RD estimates on peers' selectivity for crossing the threshold for 1st preference

	(1) Selectivity	(2) Selectivity	(3) Selectivity	(4) Selectivity	(5) Selectivity
RD estimator	0.197***	0.211***	0.217***	0.217***	0.166***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.007)
Obs.	84,770	84,770	84,770	84,770	15,011
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2013 sample of students at programs with excess demand. The outcome variable corresponds to the average test score of the students enrolled in the same program. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.24: RD estimates on enrollment for crossing the threshold for 1st preference

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr	(4) Grad by 8yr	(5) Grad by 8yr
RD estimator	0.065***	0.068***	0.067***	0.063***	0.071***
	(0.009)	(0.008)	(0.007)	(0.006)	(0.014)
Obs.	90,205	90,205	90,205	90,205	15,623
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2013 sample of students at programs with excess demand. The outcome variable indicates if the students graduate from the admission program 8 years after the admission process. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.25: RD estimates on enrollment for crossing the threshold for 1st preference

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr	(4) Grad by 8yr	(5) Grad by 8yr
RD estimator	-0.006	-0.006	-0.012	-0.021***	0.002
	(0.010)	(0.009)	(0.008)	(0.006)	(0.015)
Obs.	90,205	90,205	90,205	90,205	15,623
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2013 sample of students at programs with excess demand. The outcome variable indicates if the students graduate from any program 8 years after the admission process. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.26: RD estimate on enrollment for crossing threshold for 1st preference for boost students

	(1) enroll	(2) enroll	(3) enroll	(4) enroll	(5) enroll
RD estimator	0.199***	0.199***	0.195***	0.182***	0.193***
	(0.012)	(0.011)	(0.009)	(0.007)	(0.013)
Obs.	47,584	47,584	47,584	47,584	15,490
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2012 sample of students at programs with excess demand and a boost score larger than 5. The outcome variable indicates if the students graduate from any program 8 years after the admission process. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.27: RD estimate on graduation from the admission offer for crossing the threshold for 1st preference for boost students

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr	(4) Grad by 8yr	(5) Grad by 8yr
RD estimator	0.096***	0.095***	0.099***	0.082***	0.095***
	(0.014)	(0.012)	(0.010)	(0.008)	(0.015)
Obs.	47,584	47,584	47,584	47,584	15,490
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2012 sample of students at programs with excess demand and a boost score larger than 5. The outcome variable indicates if the students graduate from any program 8 years after the admission process. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.28: RD estimate on college graduation for crossing the threshold for 1st preference for boost students

	(1) Grad in 8yr	(2) Grad in 8yr	(3) Grad in 8yr	(4) Grad in 8yr	(5) Grad in 8yr
RD estimator	0.001	-0.011	-0.011	-0.027***	-0.001
	(0.014)	(0.013)	(0.011)	(0.009)	(0.015)
Obs.	47,584	47,584	47,584	47,584	15,490
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: 2012 sample of students at programs with excess demand and a boost score larger than 5. The outcome variable indicates if the students graduate from any program 8 years after the admission process. The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Robust standard errors are in parentheses.

Table B.29: Diff-in-diff estimates on enrollment at admission offer

	(1) Enroll	(2) Enroll	(3) Enroll	(4) Enroll
P-Up x after	0.189***	0.094***	0.276***	0.542***
	(0.017)	(0.028)	(0.030)	(0.036)
P-Down x after	-0.113***	-0.088***	-0.117***	-0.498***
	(0.015)	(0.023)	(0.026)	(0.039)
Obs.	137,400	61,188	116,834	140,033
Controls	✓	✓	✓	✓
	Moved 1-2	Moved 2-3	Moved 1-3	Moved 1-0

Notes: diff-in-diff estimates using the sample of students for whom the admission with and without the inclusion of the GPA^+ measurement moves them between the respective margins. Robust standard errors are in parentheses.

Table B.30: Diff-in-diff estimates of peers' performance at admission offer

	(1) Selectivity	(2) Selectivity	(3) Selectivity	(4) Selectivity
P-Up x after	0.250***	0.231***	0.254***	0.443***
	(0.009)	(0.013)	(0.015)	(0.011)
P-Down x after	-0.256***	-0.221***	-0.216***	-0.128***
	(0.010)	(0.012)	(0.016)	(0.013)
Obs.	137,341	61,163	116,783	139,990
Controls	✓	✓	✓	✓
	Moved 1-2	Moved 2-3	Moved 1-3	Moved 1-0

Notes: diff-in-diff estimates using the sample of students for whom the admission with and without the inclusion of the GPA^+ measurement moves them between the respective margins. Selectivity is measured as the average test score of the students in the admission program; for students without any admission, selectivity is measured as the average test score of students without any admission. Robust standard errors are in parentheses.

Table B.31: Diff-in-diff estimates for graduation from admission offer

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr	(4) Grad by 8yr
P-Up x after	0.077***	0.059**	0.160***	0.181***
	(0.019)	(0.029)	(0.033)	(0.029)
P-Down x after	-0.073***	-0.042	-0.089**	-0.240***
	(0.020)	(0.027)	(0.035)	(0.030)
Obs.	137,400	61,188	116,834	140,033
Controls	✓	✓	✓	✓
	Moved 1-2	Moved 2-3	Moved 1-3	Moved 1-0

Notes: diff-in-diff estimates using the sample of students for whom the admission with and without the inclusion of the GPA^+ measurement moves them between the respective margins. Robust standard errors are in parentheses.

Table B.32: Diff-in-diff estimates for college graduation

	(1) Grad by 8yr	(2) Grad by 8yr	(3) Grad by 8yr	(4) Grad by 8yr
Pulled-Up x after	-0.028	0.028	0.033	0.010
	(0.020)	(0.031)	(0.036)	(0.038)
Pushed-Down x after	-0.019	-0.005	-0.024	-0.111***
	(0.021)	(0.029)	(0.037)	(0.037)
Obs.	137,400	61,188	116,834	140,033
Controls	✓	✓	✓	✓
	Moved 1-2	Moved 2-3	Moved 1-3	Moved 1-0

Notes: diff-in-diff estimates using the sample of students for whom the admission with and without the inclusion of the GPA^+ measurement moves them between the respective margins. Robust standard errors are in parentheses.

Table B.33: RD estimates on enrollment for crossing threshold for 1st preference for pulled-up students

	(1) Enroll	(2) Enroll	(3) Enroll	(4) Enroll	(5) Enroll
RD estimator	0.137***	0.141***	0.122***	0.117***	0.156***
	(0.027)	(0.023)	(0.020)	(0.017)	(0.024)
Obs.	18,375	18,375	18,375	18,375	4,588
Controls	✓	✓	✓	✓	✓
Poly	5	4	3	2	1

Notes: The sample of analysis only considers students with admission at their 1st or 2nd preference (margin of treatment 1-2) and with the admission simulated without the reform at 2nd preference (therefore threshold crossing mostly due to the boost)in 2013. The sample restriction attempt to capture the effect of threshold crossing for pulled-up students by comparing them with very similar controls (non-crossing but similar score). The running variable is the application score in the most preferred choice minus the cutoff score (ex-post defined by the application of the last person admitted at that program based on the vacancies) at that program. Columns 1-4 use the entire sample. Column 5 estimates the effect with a bandwidth of 20 points (optimal bandwidth range between 10 and 20). Specifications allow for different slopes on each side of the cutoff. Robust standard errors are in parentheses.

Appendix C

Appendix of Gender bias in college admissions based on test scores: evidence and policy recommendations

C.1 Additional figures and tables

Table C.1: Translation of gender differences in graduation into applications score points

Application score equivalence to graduation difference

	(1) Graduation	(2) Application score
Female	0.065***	
	(0.009)	
Graduation		29.671***
		(0.394)
Constant	0.372***	600.737***
	(0.006)	(0.277)
Program Fixed Effect	Yes	No
Cutoff margin	Yes	No
Observations	10497	112430
Application score equivalence	1.93	

Notes: Column 1 of the table shows the difference in graduation rates between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application scores and graduation for all the students enrolled between 2012 and 2013 in programs with at least one male and one female at the margin. We use this relationship to translate the differences in graduation into differences in application scores presented at the bottom of the table.

Table C.2: Translation of gender differences in GPA into applications score points

	(1) Grad
Female	0.075***
	(0.003)
Wage job	-0.047***
	(0.006)
Urban housing	0.035***
	(0.003)
Private health	0.020***
	(0.003)
Mother with college	0.001
	(0.004)
Father with college	0.003
	(0.004)
Program Fixed Effect	Yes
Cutoff margin	No
Observations	120,503

Notes: Column 1 of the table shows the differences in 1st year GPA between females and males around the cutoff - 10% of students - considering only programs with at least one male and one female in that margin and with year and program fixed effect. Column 2 presents the relationship between application score and 1st year GPA for all the students enrolled between 2013 and 2017 in programs with at least one male and one female at the margin. We use this relationship to translate the differences in 1st year GPA into differences in application scores presented at the bottom of the table.