NBER WORKING PAPER SERIES

HOW FAR CAN INCLUSION GO? THE LONG-TERM IMPACTS OF PREFERENTIAL COLLEGE ADMISSIONS

Michela Carlana
Enrico Miglino
Michela M. Tincani

Working Paper 32525
http://www.nber.org/papers/w32525

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

We thank seminar participants at PUC, Tufts, LUISS, IFS, Queen Mary, UCL, Stockholm University, the VATT Institute and at various workshops for helpful comments. Sara Chiuri and Sofia Betania Sierra Vasquez provided excellent research assistance. AER RCT Registry ID: AEARCTR-0002288. The project has obtained IRB approval from Harvard University (IRB23-0805) and UCL (10515/002). We acknowledge support from HKS Malcolm Wiener Center for Social Policy. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Far Can Inclusion Go? The Long-term Impacts of Preferential College Admissions
Michela Carlana, Enrico Miglino, and Michela M. Tincani
NBER Working Paper No. 32525
May 2024
JEL No. I24

ABSTRACT

Affirmative action and preferential admission policies play a crucial role in fostering social mobility by bolstering the prospects of disadvantaged groups. In this paper, we analyze the long-term effects of a Chilean policy (PACE) that targets students in underprivileged schools, offering guaranteed admission to selective colleges to those graduating in the top 15 percent of their high school class. Leveraging both the randomized expansion of PACE and the admission discontinuity, our analysis reveals that PACE yields positive labor market effects for the average targeted student, especially women, driven by the selectivity of the attended colleges. However, for marginally eligible students, higher dropout rates and negative labor market outcomes emerge, suggesting PACE may induce a mismatch between their skills and the academic rigor of selective programs. Finally, we find that students in the bottom 85 percent of their schools experience positive effects on labor market outcomes. We identify equilibrium effects on local labor markets as a potential mechanism. The results suggest that there is a limit to how far preferential admissions can go while delivering on their promises.

Michela Carlana
Harvard Kennedy School
79 JFK Street
Taubman Bldg
Cambridge, MA 02138
and IZA
and also NBER
michela_carlana@hks.harvard.edu

Enrico Miglino
Bank of Italy
miglinoenrico@yahoo.it

Michela M. Tincani
University College London
Gower Street
London - WC1E 6BT
UK
m.tincani@ucl.ac.uk

A randomized controlled trials registry entry is available at AEARCTR-0002288

How Far Can Inclusion Go? The Long-term Impacts of Preferential College Admissions *

Michela Carlana[†], Enrico Miglino[‡], and Michela M. Tincani[§]

May 2024

Abstract

Affirmative action and preferential admission policies play a crucial role in fostering social mobility by bolstering the prospects of disadvantaged groups. In this paper, we analyze the long-term effects of a Chilean policy (PACE) that targets students in underprivileged schools, offering guaranteed admission to selective colleges to those graduating in the top 15 percent of their high school class. Leveraging both the randomized expansion of PACE and the admission discontinuity, our analysis reveals that PACE yields positive labor market effects for the average targeted student, especially women, driven by the selectivity of the attended colleges. However, for marginally eligible students, higher dropout rates and negative labor market outcomes emerge, suggesting PACE may induce a mismatch between their skills and the academic rigor of selective programs. Finally, we find that students in the bottom 85 percent of their schools experience positive effects on labor market outcomes. We identify equilibrium effects on local labor markets as a potential mechanism. The results suggest that there is a limit to how far preferential admissions can go while delivering on their promises.

1 Introduction

Despite efforts to promote equal access to higher education, college degrees remain disproportionately attained by students from advantaged backgrounds (Chetty et al., 2020). Affirmative action and preferential admission policies tackle such inequality by enhancing admission chances for disadvantaged groups. Most research reveals substantial benefits to

*We thank seminar participants at PUC, Tufts, LUISS, IFS, Queen Mary, UCL, Stockholm University, the VATT Institute and at various workshops for helpful comments. Sara Chiuri and Sofia Betania Sierra Vasquez provided excellent research assistance. AER RCT Registry ID: AEARCTR-0002288. The project has obtained IRB approval from Harvard University (IRB23-0805) and UCL (10515/002). We acknowledge support from HKS Malcolm Wiener Center for Social Policy.

[†]Harvard Kennedy School, NBER, CEPR, and LEAP (e-mail: michela_carlana@hks.harvard.edu).

[‡]Bank of Italy (e-mail: Enrico.Miglino@esterni.bancaditalia.it).

[§]University College London, IFS, CEPR, CESifo, and LEAP (e-mail: m.tincani@ucl.ac.uk).

well-prepared, underrepresented students admitted to selective colleges (Bleemer, 2022; Black, Denning, and Rothstein, 2023; Otero, Barahona, and Dobbin, 2023). However, the benefits for disadvantaged students with lower academic preparation remain an open question, as much of the prior work has centered on relatively high-achieving underrepresented applicants. This question has garnered significant relevance following the recent U.S. Supreme Court's ban on race-based admissions. Institutions are pivoting towards admission preferences targeted at students from low-income families. These students have fewer opportunities to enhance their human capital during childhood and adolescence (Cunha and Heckman, 2007), with negative consequences for their achievement by the end of compulsory schooling (Belley and Lochner, 2007; Bailey and Dynarski, 2011; Chetty et al., 2020; Hanushek et al., 2022). This paper investigates the impacts in the labor market of extending preferential admissions to disadvantaged students, substantially underprepared compared to their peers.

We analyze the effects of a Chilean affirmative action policy called PACE (*Programa de Acompañamiento y Acceso Efectivo a la Educación Superior*), which targets students in disadvantaged schools and offers those graduating in the top 15% of their high school guaranteed admissions to selective colleges. Although top graduates within their schools, the high school standardized test scores of targeted students are 0.9 standard deviations below those of regular college entrants. For those just around the top 15% admission cutoff, this difference reaches 1.1 standard deviations.

We evaluate the impacts of the policy on students labor market outcomes and educational attainment. We use two research designs. First, we exploit the randomized expansion of the program that occurred in 2016 to identify the effect of being in a school that participates in PACE. Our randomized control trial analysis (RCT) considers separately the impacts on the top-15 percent of students, whom the policy aimed to attract to college, and on the bottom-85 percent of students. Second, we exploit the discontinuity in the admission rule to isolate the effect of receiving a preferential admission offer. The regression discontinuity approach (RDD) focuses on students in schools participating in the PACE program since 2014, whose grades are close to the preferential admission cutoff. The two approaches provide complementary evidence. The RCT allows us to estimate the effect of introducing affirmative action in schools for both top-performing students and their schoolmates, combining the effect of preferential admissions with any other changes, such as to the incentives to study while in school (Tincani, Kosse, and Miglino, 2024) or the orientation classes provided by PACE to all students in participating schools. The RDD narrows down the impact exclusively of preferential admission by comparing students just above and just below the preferential admission cutoff. In terms of populations of interest, the RCT provides the treatment effect throughout the baseline ability distribution, while the RDD provides local effects at the eligibility margin.

We construct a unique panel dataset merging together administrative data on education and labor market outcomes up to six years after leaving high school. For the cohort graduating in 2016, we additionally link survey data on expectations about future college performance that we collected at the end of high school. Furthermore, we create a separate dataset linking administrative data on education and labor market outcomes for older cohorts of students who graduated from the same schools as our study subjects between 2006 and 2010. This dataset allows us to implement the surrogate index method (Athey, Chetty, Imbens, and Kang, 2019) to predict program impacts on labor market outcomes up to fifteen years after high school, allowing us to provide some of the longest-term estimates of preferential admission impacts.

First, we provide evidence on the effects of PACE on targeted students, the top 15% of their school, including those around the preferential admission cutoff that the program aimed to bring to college. A long-standing theory, the 'mismatch hypothesis' (Sander, 2004), posits that preferential admissions targeted at low-achieving students could make them worse off by inducing them to enroll in selective colleges only to later drop out, lowering their lifelong earnings. While existing evidence has so far mostly disputed this hypothesis (Black, Denning, and Rothstein, 2023; Bleemer, 2022, 2021; Otero, Barahona, and Dobbin, 2023), PACE targets a population of students with substantially lower pre-college achievement than those regularly admitted to college, and than those targeted by similar programs in other countries. Whether mismatch can occur in such a population remains an open empirical question.

In the RCT analysis, we do not detect significant impacts on labor force participation or earnings for top 15% students in the first five years after high school. Using the surrogate index method, we predict positive effects on yearly earnings in the long term, with an increase by 3.9% in the yearly earnings between 11 and 15 years after high-school graduation. Consistent with the null impacts on labor force participation, PACE did not increase higher education attendance on average in this group. While enrollment in higher education increased substantially the first year after high school (+6.2 percentage points (p.p.), a 10% increase compared to the mean of the control group), PACE did not alter the likelihood of enrolling within six years of high school. The positive predictions on long-term yearly earnings are driven by an increase in graduation from selective colleges (+ 6.2 p.p., a 22% increase), with a similar decrease in graduation from vocational and non-selective institutions.The long-term positive effect on the labor market is fully driven by women, with an yearly earning increase of 9% in the long term, likely driven by their higher graduation from selective colleges (+10.6 p.p., or 36% increase). The effect of PACE for high-achieving men is negligible, both in the labor market and higher education. In line with previous studies, therefore, we do not find evidence of *global* mismatch among the pool of targeted students on average (Arcidiacono and Lovenheim, 2016) and more positive impacts on women (Black, Denning, and Rothstein, 2023). In contrast to previous studies, we find that the benefits are fully concentrated among women. When severely disadvantaged top-performing men are offered the opportunity to attend selective colleges, they do not increase their probability of graduating from them, limiting their potential for long-term social mobility.

The results for students just eligible for a preferential admission, however, are strikingly different compared to the average effect for all targeted students. Marginal students are systematically less likely to work when eligible for a PACE slot in the five years after high school, with a 10% decrease in the number of months worked. This effect is accompanied by a 10% yearly earning loss during this period. In the long term (years 11-15 after high school graduation), the surrogate index method predicts a positive but insignificant effect on earnings. Consistent with the negative impact on labor force participation in the short term, PACE admissions increased the likelihood of enrolling into any tertiary institution within six years of high school (4.1 p.p., or a 5% increase), suggesting that some students were induced to attend higher education instead of entering the labor market, an extensive margin effect. Preferential admissions boosted selective college degree attainment for some marginal students (+5.5 p.p.), but also increased the likelihood to enroll in tertiary education only to later drop out (+6.5 p.p.). While some of this dropout likely occurred among the new entrants, some may have occurred among those who would have obtained a higher education degree (potentially from a less selective institution) without the preferential admission.

The heterogeneity by gender offers valuable insights on these results to understand who benefits from affirmative action policies targeting very disadvantaged students. Preferential admissions boosted selective college degree attainment for some marginal students of both genders (5.5 p.p.), but they also increased the likelihood to enroll in tertiary institutions only to later drop out for some men (12.7 p.p), with negative implications for earnings, consistent with preferential admissions inducing mismatch among some marginal male students. PACE yielded positive long-term labor market outcomes for some marginal students, of both genders, and negative for others, especially men. To further investigate what is driving the negative RDD results, we assess whether information frictions preclude some students from correctly foreseeing the consequences of enrolling in a demanding college program. This is a central tenet of the 'mismatch hypothesis' (Arcidiacono et al., 2011), that has so far been difficult to directly test absent expectation data.

Using survey data we collected from students in a sample of high schools in 2016 during their final year and linked to their administrative records, we show that students, especially men, are overoptimistic about their chances of graduating from a selective college. For example, nearly 40% of college entrants who were certain they would graduate from a selective college had instead dropped out during the following six years. We show supporting evidence that the negative impacts on the higher education performance of marginal students are concentrated among students who, in high school, had above-median levels of overconfidence in their belief that accepting admission to a selective college would lead to graduation. In contrast, and in line with the revealed preference theory, students who held more realistic expectations faced no negative effects when offered additional opportunities to attend college. This evidence is entirely consistent with the theory that affirmative action and preferential admissions can lead to mismatch effects in the presence of information frictions for low-achieving and overoptimistic students (Arcidiacono et al., 2011).

Next, we discuss the PACE effects on untargeted students, those belonging to the bottom 85% of their school in terms of GPA at the experiment's baseline. Preferential admission policies can influence schoolmates of targeted students through at least three channels. Exposure to peers headed for college could lead them to pursue higher education at higher rates through a desire to conform (Golightly, 2019; Fernández, 2021; Anelli and Peri, 2019). Not being awarded a preferential admission could act as a negative signal on own academic ability and constrain educational aspirations, leading students to invest in their labor market prospects instead (Genicot and Ray, 2017; La Ferrara, 2019). Finally, admission policies could generate equilibrium effects in local labor markets by increasing the supply of college attendants (Moretti, 2004).

In the RCT analysis for bottom 85% students, we find sizeable positive impacts on labor force participation (+7% more months worked) and earnings (+8%) in the first five years after high school, with the initial earning boost predicted to persist over time, although decreasing in magnitude and not precisely estimated. We do not find any effects on educational attainment in this group, ruling out the peer spillovers channel. Examining patterns of entrance exam re-taking over subsequent rounds, we find no evidence that the increased labor force participation came at the cost of a disengagement with the higher education sector, as would be expected if students interpreted the absence of preferential admission as a negative signal constraining their educational aspirations. Instead, we find support for equilibrium effects in local labor markets. When we compare students in PACE and non-PACE schools that are located in the same labor market through the inclusion of labor market fixed effects, the positive effects of being in a PACE school on labor market outcomes in the short term vanish. The positive effects we detected, therefore, could be due to changes occurring in the labor markets hosting treated schools. Since targeted students from treated schools attend higher education for longer than their counterparts in control schools during the first five years after high school, a plausible mechanism is that there are more job vacancies and less competition for jobs in the local labor markets hosting treated schools.[1] While such imbalances in vacancies across labor markets fluctuate over the first five years, in frictional labor markets they can generate constant employment differences like those we documented.

The paper contributes several novel findings to the literature on college admissions. First, it extends the literature on preferential admissions based on socioeconomic disadvantage, providing labor market impact estimates for a population of students with lower pre-college achievement than the populations targeted by similar programs in other countries (Black, Denning, and Rothstein, 2023; Bleemer, 2021). This allows us to provide new empirical insights to the debate over mismatch in affirmative action (Sander, 2004; Arcidiacono and Lovenheim, 2016). Unlike previous studies, we find evidence consistent with the mismatch hypothesis for some marginally eligible men, but not for the average students targeted by the policy. Substantially relaxing admission requirements may attract into selective colleges men who would have otherwise entered the labor market. As several preferentially admitted marginal men drop out of higher education, the short-term earning losses are not necessarily offset by higher future wages.[2,3] Such short-term earning losses have not been documented in preferential admission contexts that reach more academically-able students, for whom a more common counterfactual choice would be to enter a less selective college.[4]

Our evidence contributes in several ways to the affirmative action literature. First, it demonstrates that mismatch is a local phenomenon even in contexts in which admission preferences are substantially extended. The results also speak to the external validity of analyses based on local estimates around admission cutoffs. In the context of affirmative action for higher education, local effects may not generalize away from admission cutoffs as well as in other education contexts (Angrist and Rokkanen, 2015).

Second, we provide new evidence on the equilibrium impacts that affirmative action policies can have on untargeted *disadvantaged* students. Prior studies have mostly focused on examining the impacts on untargeted advantaged students, either those displaced

[1]While the treatment did not increase the likelihood of entering higher education for top-performing students, targeted students in treated schools enter higher education earlier, and attend longer programs. As a result, in the first year after high school, and then again two years later, higher education attendance is larger among targeted students from treated schools than among their counterparts in control schools.

[2]Using a generalized Roy framework applied to NLSY79 data from the United States, Heckman, Humphries, and Veramendi (2018) estimate insignificant effects of college graduation on the wages of students with below-median academic ability in the U.S. They conjecture that the short-term earning losses from entering college instead of the labor market for such students are not offset by higher future wages, but their estimates are noisy because few students in this group enter college absent preferential admission programs.

[3]Some marginally admitted men may have obtained vocational degrees absent preferential admissions. By foregoing tertiary degrees, these men may experience long-term wage losses.

[4]Analyzing the Californian top percent plan, Bleemer (2021) finds only intensive-margin effects on college quality. Analyzing the Texas top percent plan, Black, Denning, and Rothstein (2023) find extensive margin effects on the likelihood of attending higher education at all, but do not study short-term earning impacts.

from college opportunities (Black, Denning, and Rothstein, 2023; Bleemer, 2021; Otero, Barahona, and Dobbin, 2023), or college incumbents (Machado, Reyes, and Riehl, 2023).[5] Our evidence of positive equilibrium effects on the labor market outcomes of untargeted disadvantaged students demonstrate that a comprehensive cost-benefit analysis of the impacts of these policies should consider their potential to affect this group of untargeted students.[6]

2 Institutional background

2.1 Higher education in Chile

Chile's higher education system encompasses three distinct types of institutions: selective colleges, non-selective colleges, and vocational institutions. These categories contribute to higher education enrollment in varying proportions, with selective colleges accounting for 41%, non-selective colleges representing 8%, and vocational institutions constituting the majority at 51% in 2018.

Selective colleges include the thirty-nine that participate in the centralized admission system (SUA, *Sistema Único de Admisión*). Renowned for their academic focus, these institutions provide comprehensive four to six-year programs. To enroll, students must take a standardized college entrance exam, which was called PSU (*Prueba de Selección Universitaria*) up to 2020. Students submit an application on the centralized platform, and an allocation mechanism matches applicants to programs. Like in most industrialized countries, enrollment in selective colleges is highly unequal across socioeconomic lines (Figure A1).

2.2 The PACE program

PACE (*Programa de Acompañamiento y Acceso Efectivo a la Educación Superior*) is a preferential admission policy introduced to increase admissions to selective colleges among disadvantaged students. The program is underpinned by an agreement between the gov-

[5] An exception is Golightly (2019), who examines impacts on the education outcomes of students in schools targeted by the Texas Top Ten policy but who are not top-decile students.

[6] This paper extends Tincani, Kosse, and Miglino (2024), who focus on the impacts of PACE on pre-college outcomes and education outcomes in the first five years after high school. Tincani, Kosse, and Miglino (2024) combine the RCT with a structural model of education choices to rationalize PACE impacts on pre-college effort, and do not exploit the discontinuity to examine admission impacts among marginally eligible students. Therefore, the two central contributions of this paper — identifying how much admission requirements can be extended while still benefiting the long-term outcomes of targeted students, and providing evidence of spillover effects on the labor market outcomes of untargeted disadvantaged students — are entirely novel.

ernment and the selective colleges, which commit to offering a list of reserved seats to PACE students.

The PACE program targets high schools selected by the government based on a school-level vulnerability index (*Índice de Vulnerabilidad Escolar*). It was first introduced in disadvantaged Chilean high schools in 2014, and later expanded, as can be seen from the map in Figure 1 and the dynamic graphic available here. The first high schools to participate in PACE in 2014 were located in the poorest regions of south-central Chile where most indigenous communities live (e.g. Auracanía) and in the poorest neighborhoods of the southwestern Santiago metropolitan area. The program expanded gradually over time, incorporating schools selected from the poorest areas of the country. When a school first enters the PACE program, the preferential admission criteria apply only to the cohort in 11^{th} grade; in subsequent years, they apply also to the new cohorts entering 11^{th} grade.

After the first two years of program implementation, the government identified 218 high schools that were not yet PACE schools, but that met the eligibility criteria for entering PACE in 2016 given students' socioeconomic status.[7] Following advice from our research team and using a randomization code written by PNUD Chile (United Nations Development Program), it randomly selected 64 of the 218 eligible schools to receive the PACE treatment. Figure A2 shows the geographic distribution of the experimental schools. The randomized expansion concerned a single cohort of students: those starting 11^{th} grade in March 2016. Before starting the school year, students who were enrolled in high schools randomly selected to be treated were informed their school was in the PACE program. The announcement was made after the school enrollment deadline; thus, there is no strategic selection into high schools (Tincani, Kosse, and Miglino, 2024).[8]

Students in high schools participating in PACE can apply to a selective college through the regular channel, like any other student in the country. Moreover, they are offered a guaranteed admission to a selective college if they satisfy three conditions:

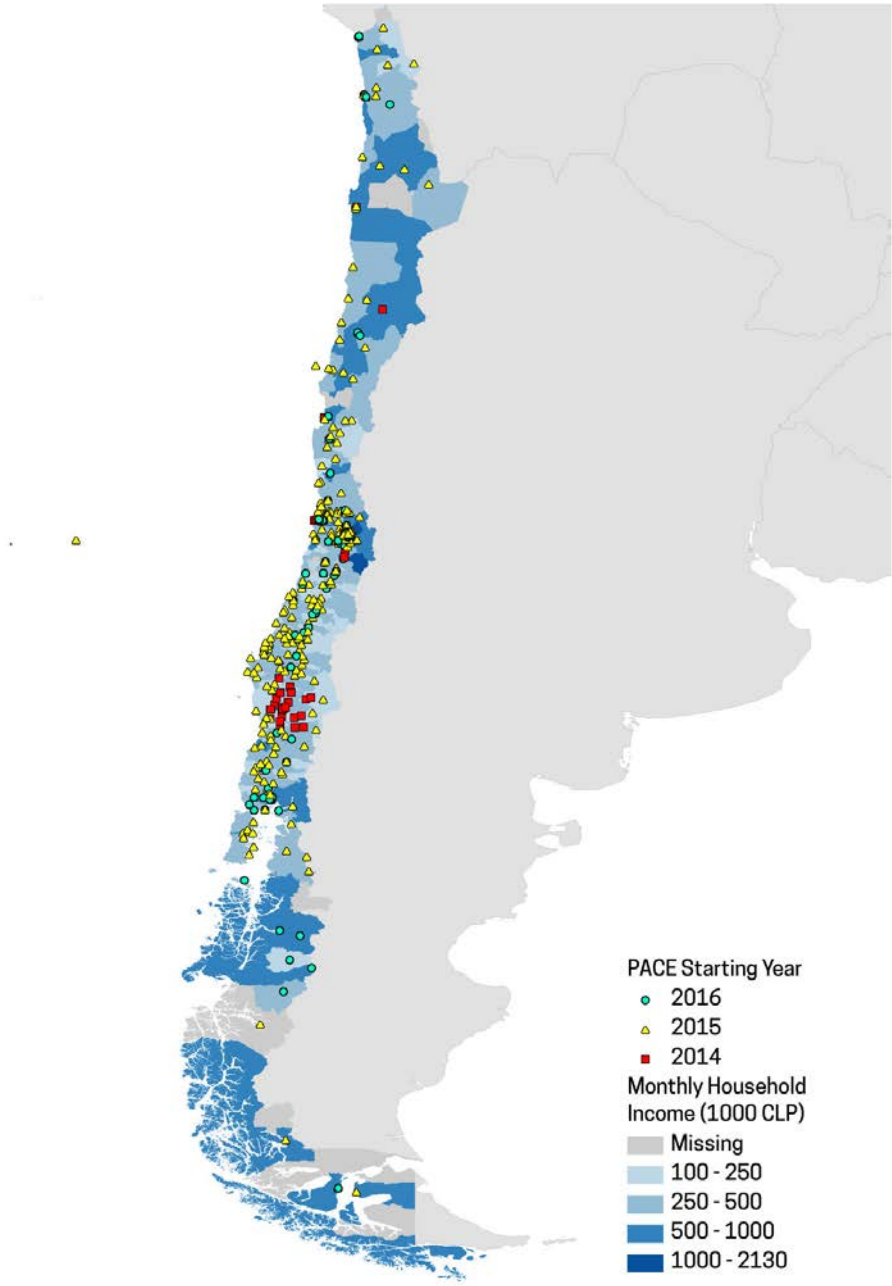
1. Being above the 85^{th} percentile of the high school GPA distribution or having obtained a GPA score equal to or greater than a national cutoff.[9] This feature makes PACE a 'percent plan' program.

[7]The government initially identified 221 disadvantaged high schools as a potential target of the program expansion, but 3 of them had no students in the cohort starting 11^{th} grade in 2016, the target cohort, and were eliminated from the list.

[8]Grades in the first two high school years (9 and 10) were already determined when students in treated schools were informed they were in a PACE school, affecting at least in part the GPA used as eligibility cutoff for PACE, called PRN (*Puntaje Ranking de Notas*).

[9]The national cutoff is binding for only 13% of students in our study population. The GPA score used to rank students, called PRN (*Puntaje Ranking de Notas*), is calculated from the grade point average in grades 9 to 12 adjusted to account for the school context. The Pearson's correlation coefficient between the unadjusted four-year grade point average and the PRN score is 97.44%. The national cutoff changed over the years: it was 710 points in 2016, 705 points in 2017, and 703 points in 2018.

Figure 1: Geographic distribution of PACE high schools.



Notes: This figure shows the heatmap of Chile in terms of average monthly household income (1000 Chilean Pesos) and the geographic distribution of high schools that entered the PACE program between 2014 and 2016, the program cohorts included in this study.

2. Having attended a school that participates in the PACE program in grades 11 and 12, and having attended the light-touch orientation classes offered to all students in PACE schools during grades 11 and 12.

3. Having taken the entrance exam for selective colleges, even though the score does not affect the likelihood of obtaining a PACE admission.

Unlike other percent plans, such as those implemented in Texas and California (Horn and Flores, 2003, 2015), there are no additional required coursework to be eligible. However, up to 2020, the PACE guaranteed admission to selective colleges could be used only in the year immediately after graduating from high school, potentially inducing students to start college right after high school.

The selective college seats offered through PACE are similar to those offered through the regular admission channel in terms of major, selectivity (i.e. the average entrance exam of regular entrants), and distance from the student's home (Tincani, Kosse, and Miglino, 2024). PACE is still relatively small relative to the size of the college sector. In 2018, only 2.18% of all college enrollments were through PACE, distinguishing this program from large-scale affirmative action policies that induce large inflows of quota students into college (Machado, Reyes, and Riehl, 2023). This program feature implies that grading and teaching standards are not expected to adjust to the influx of preferentially-admitted students. Therefore, if the program leads students to enroll in programs for which they are under-prepared, we expect to see evidence of dropout. Given the program's size, preferential college seats are supernumerary: they do not replace regular seats but are offered in addition to them. Finally, a student could obtain both a PACE and a regular admission: if she or he did not accept a PACE admission, that PACE seat remained vacant. Appendix D describes the seat allocation mechanism.

While PACE seats are not tied to financial aid, practically all students eligible for PACE seats are also eligible for tuition fee waivers. This is because all students coming from families in the bottom 60% of the income distribution are eligible for such waivers.

3 Data

3.1 Data sources

We use data from several sources, summarized in Table 1, which we were able to merge with each other at the individual student level through a collaboration with the Chilean Ministries of Finance and of Education.[10]

From school enrollment registries and the SIMCE (*Sistema de Medición de la Calidad de la Educación*) and SEP (*Subvención Escolar Preferencial*) datasets from the Ministry of Education we obtained information on high school students' academic performance, background characteristics, and school characteristics. We complemented these data with survey data we collected in 2017 in some of the schools participating in the randomized experiment, giving us information on students' expectations about their future persistence in higher education. We then observe actual higher education enrollments, graduation and

[10]These data will be complemented with tax records obtained from the Internal Revenues Services.

Table 1: Overview of Data Sources

DATASET	VARIABLES	TIMING	SOURCE
1. *SIMCE*	Achievement test scores, background characteristics	Grade 10	Admin
2. *SEP*	Very-low-SES classification (*Prioritario* student)	Grade 10	Admin
3. School records	High-school enrollment, school characteristics	Grades 9-12	Admin
4. Student survey	Expectations	Grade 12	Primary
5. Higher education records	Engagement with centralized application system, enrollments, graduation and persistence in any higher education institution, program selectivity	Up to 6 years after high school	Admin
6. PACE program records	Allocation of PACE seats in selective colleges, enrollments, graduation and persistence via PACE channel, program type, selectivity, and major	Up to 6 years after high school graduation	Admin
7. Unemployment insurance	Months worked, earnings, occupation	Up to 5 (actual) and 15 (predicted) years after high school	Admin

persistence, through the regular and the PACE channels, obtained from higher education records from the Higher Education Division of the Ministry of Education. We observe months worked, earnings and sector obtained from the unemployment insurance dataset (SC, *Seguro de Cesantiía*), housed at the Ministry of Finance.[11]

3.2 Samples and descriptive statistics

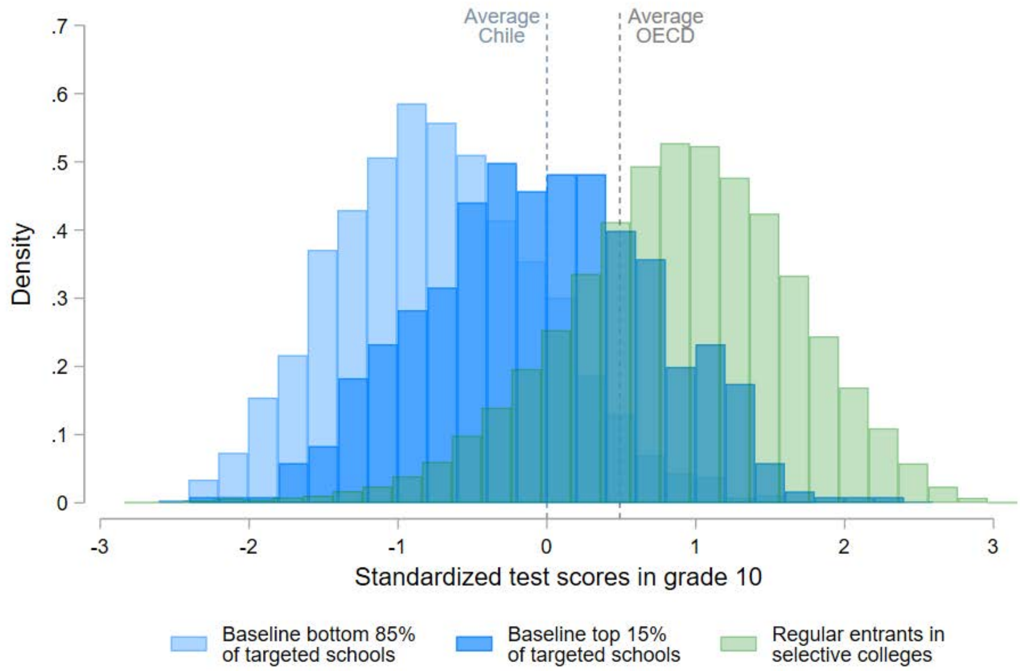
To investigate the overall effect of PACE, we exploit three main samples: (i) *RCT Top 15% sample*: all students in the top 15% of high school GPA, as measured by the average GPA in grade 9 and 10, in all schools of the experimental sample, (ii) *RCT Bottom 85% sample*: all students in the bottom 85% of high school GPA, measured in grade 9 and 10, in all schools of the experimental sample, (iii) *RDD Sample*: marginal students around the top 15% cutoff in PACE schools, calculated in grades 9 to 12 as by eligibility cutoff of the policy.[12] For the RCT we calculate the GPA only using data in grade 9 and 10, before the policy announcement at school level, to avoid potential treatment-dependent selection

[11]The SC dataset does not include information on self-employment and the public sector, a data limitation shared by other studies of preferential admission impacts (Bleemer, 2022). The SC dataset also censors very high earnings, but this is not likely to be problematic in our sample of very disadvantaged students whose earnings are typically below the censoring cutoff (only 0.18% of observations in our sample are censored).

[12]The adjusted GPA in grade 9 to 12 used to rank students, i.e. the PRN score, is released to researchers only for the sample of students who registered to take the entrance exam to selective universities (PSU).

into the top 15% in grade 11 and 12. The RCT samples include the cohort of students in 11^{th} grade in 2016; the RDD sample includes the cohorts in 11^{th} grade between 2014 and 2016.

Figure 2: Academic preparation: high school test scores of regular college entrants vs students in PACE schools



Notes: This figure shows the distribution of grade 10 test scores among students in PACE schools and among regular college entrants. Those in PACE schools are divided into the sub-samples belonging to the bottom 85% and the top 15% of their school based on GPA in grades 9 and 10. The test score is standardized to have mean zero and variance one in the population of ten graders. The average 10^{th} grade standardized test scores in the OECD are calculated by re-scaling the PISA scores following the procedure described in the supplementary material of Tincani, Kosse, and Miglino (2024).

The students in our samples exhibit low academic readiness: while regular entrants in selective colleges display an average baseline test score in grade 10 at the 79^{th} percentile of the nationwide ability distribution, students within the top 15% of PACE schools find themselves at the 48^{th} percentile. On average, marginal students fall around the 41^{st} percentile, while the bottom 85% of students possess an average ability corresponding to the 26^{th} percentile of the nationwide test score distribution in grade 10. Figure 2 illustrates this pattern, revealing that the top 15% and bottom 85% score 0.88 and 1.56 standard deviations below regular entrants in selective colleges, respectively. Marginal students around the cutoff vary in their baseline distribution due to differences in academic readiness across schools in the sample, but on average they score 0.21 standard deviations below the top 15% (Figure A3).

Table 2 presents descriptive statistics of the three study samples beyond the baseline test scores. Female students are over-represented among the top students in the PACE schools, especially around the cutoff margin. Around 60 percent of students in all samples are categorised by the government as having very low socioeconomic status and around two in three attended a vocational high school track. Students in the bottom 85% have lower ability as measured not only by the standardized test score, but also grade failure.

Table 2: Baseline characteristics, experimental and RD samples

	Top 15%			RD			Bottom 85%		
	(1) Obs.	(2) Mean	(3) S.D.	(4) Obs.	(5) Mean	(6) S.D.	(7) Obs.	(8) Mean	(9) S.D.
Female	2437	0.56	0.50	15103	0.59	0.49	11916	0.47	0.50
Age	2437	16.32	0.64	15103	16.34	0.65	11916	16.60	0.79
Very low SES	2437	0.59	0.49	15103	0.61	0.49	11916	0.61	0.49
Mother's education	1914	9.68	3.10	9856	9.98	3.16	7754	9.53	3.14
Father's education	1795	9.50	3.15	9322	9.79	3.25	7362	9.32	3.26
Family income	1919	284.05	195.67	9854	310.15	254.69	7782	289.21	214.89
SIMCE	2432	-0.01	0.82	14972	-0.21	0.77	11875	-0.69	0.71
Never failed	2437	0.94	0.24	15103	0.92	0.27	11916	0.81	0.39
Santiago	2437	0.17	0.37	15103	0.22	0.41	11916	0.16	0.37
Rural	2437	0.04	0.19	15103	0.05	0.22	11916	0.03	0.18
Academic track	2437	0.31	0.46	15103	0.37	0.48	11916	0.26	0.44

Notes: This table reports the number of observations, mean and standard deviations for pre-determined variables for the experimental sample, split in top 15% and bottom 85% according to the baseline high school GPA ranking, and for the RD sample within the optimal bandwidth for the first-stage regression estimated using the method by Calonico, Cattaneo, and Titiunik (2014). SIMCE is a standardized achievement test taken in 10^{th} grade. Age and education are in years. Family income is the monthly family income in 1000 Chilean pesos. p<0.01, ** p<0.05, * p<0.10.

Next, we provide an overview of labor market and higher education outcomes, referring the interested reader to Appendix C.1 for a detailed description of variable construction. For each of our three samples, we measure the number of months worked and earnings in each year after graduation. In Table A1, we report the summary statistics of the average yearly earnings and number of months worked in the short term (1-5 years after high school graduation), in the medium term (6-10 years after high school graduation), and in the long term (11-15 years after high school graduation). In the short term, students in the top 15% of PACE schools worked for around 3.3 months and earned 1.7 million pesos, almost 2,000 USD. Students in the bottom 85% worked and earned slightly more (4.6 months and 2.4 million pesos) due to their lower participation in tertiary education.

We predict medium-term and long-term labor market outcomes using the surrogate index method (Athey et al., 2019), which we describe in detail in Appendix C.2. The method requires access to data on older individuals who are comparable to those under study but for whom longer-term outcomes are observed. To implement it, we obtained data on the population of students who graduated from the same schools used in the analysis, between 2006 and 2010. We merged their education and labor market records, observing outcomes up until fifteen years after high school graduation. Appendix Table A2 provides a summary of the baseline characteristics and outcomes of the sample used to construct the predictions. As expected, these students display a very similar background to the students in our main study samples. Using the method, the predicted earnings of the students in our study samples by year fifteen post high school are substantially higher for students in the top of the GPA distribution as they are more likely to graduate from college and get highly paid jobs. Table A1 shows that the predicted yearly earnings in the long term are 15 percent higher for students in the top 15% compared to the bottom 85% (5.8 million pesos vs. 4.2 million pesos). Predicted earnings in the long term (11-15 years) decline with respect to the medium term (6-10 years) due to a 34% drop in earnings for women driven by a 45% drop in months employed, whereas men's earnings are virtually the same in the two periods.

The educational outcome variables of interest include whether, in the six years after high school: (i) a student ever enrolls in higher education, (ii) a student enrolls and either graduates or is still enrolled–and, therefore, can potentially graduate–from higher education ('graduation') (iii) a student enrolls in higher education in any year after high school and later drops out. We calculate each of these three main outcome variables for selective colleges, directly targeted by PACE, and any tertiary institution to investigate potential substitution effects.

Panel B of Table A1 reports the main summary statistics of these outcomes: among the students in the top 15%, 87 percent have been enrolled in any tertiary institution and 47 percent in a selective college. As expected, the probability of ever enrolling in higher education is substantially lower for students in the bottom 85%, standing at only 67 percent in any institutions and 15 percent for selective colleges. Less than two out of three students in the top 15% are expected to graduate from tertiary education and one in three from a selective college. Only 4 out of ten students in the bottom 85% are expected to graduate from higher education and only one in ten from a selective institution. The statistics for the sample of marginal students are close to the top 15%, with a slightly lower selective college attendance, as expected.

4 Empirical methodologies

We use two research designs. First, we implement a Randomized Controlled Trial (RCT). We exploit the randomized expansion of the program that occurred in 2016 to identify the effect of being in a school that participates in PACE. Our analysis considers separately the impacts on the top 15% of students in terms of GPA, whom the policy aimed to attract to college, and the bottom 85%. Second, we implement a Regression Discontinuity Design (RDD). We exploit the discontinuity in the admission rule to isolate the effect of receiving a preferential admission offer. This approach focuses on students in PACE schools, in 11^{th} grade between 2014 and 2016, whose grades are close to the preferential admission cutoff based on the high school GPA.

The two approaches provide complementary evidence on the impacts of PACE. The RCT allows us to estimate the effect of introducing affirmative action in schools for both top-performing students and their schoolmates, combining the effect of offering preferential admissions to targeted groups with any other changes, such as to the type of light-touch orientation offered to students or to the incentives to study while in school (Tincani, Kosse, and Miglino, 2024). The RDD narrows down the impact exclusively of preferential admission by comparing students in PACE schools who are just above and just below the preferential admission cutoff. In terms of populations of interest, the RCT provides the treatment effect throughout the baseline ability distribution, while the RDD provides local effects. The RCT includes only the cohort of students in grade 11 in 2016, while the RDD includes also the two older cohorts.

4.1 Randomized Controlled Trial

Among the 218 disadvantaged high schools selected by the Chilean government as a potential target for the PACE expansion of 2016, 64 were randomly assigned to PACE. The remaining 154 schools were not offered PACE nor promised future entry—they were not contacted by Ministry officials. We estimate the following linear regression model:

$$Y_{is} = \alpha + \beta T_s + \lambda X_i + \varepsilon_{is}, \tag{1}$$

where Y_{is} is the outcome of student i in school s, T_s is a dummy variable equal to 1 if school s was randomly selected to participate in PACE and to 0 otherwise, and X_i is a vector of student i's baseline characteristics.[13] The standard errors are clustered at the high school level.

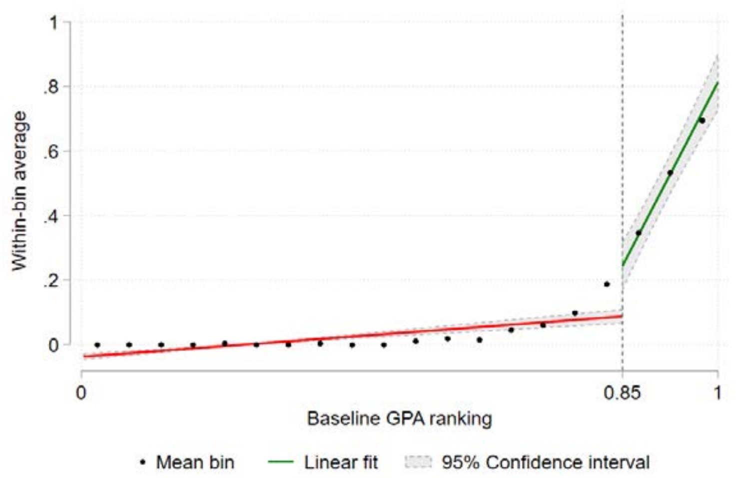
As expected, PACE admission offers are made almost exclusively to students who at baseline were in the top 15% of their PACE school based on GPA in grades 9 and 10; 54.2% of students in the top 15% and 2.7% of those in the bottom 85% of GPA received a PACE admission offer (Figure 3).[14]

As shown in Table A3, the sub-samples of students in the top 15% and bottom 85% of their school in terms of baseline high school grades are balanced across the treatment and control groups in terms of all pre-determined covariates, i.e., gender, age, socioeconomic status, mother's and father's education, family income, baseline standardized test scores,

[13] The controls are: gender, age, an indicator for whether the student is identified by the government as socioeconomically very vulnerable (*Alumno Prioritario*), baseline grade 10 SIMCE test score, whether the student never failed a grade, and the high school track (academic or vocational). We present robustness results without controls in the Appendix.

[14] The share of students receiving the offer is around 10.4% of students in the schools as some students in the top 15% may have not satisfied the other criteria to receive a PACE admission offer (e.g., they may have not taken the PSU, or attended a PACE school for both grade 11 and 12, or they may have not attended the orientation classes).

Figure 3: Preferential college admission offers.



Notes: This figure shows how the likelihood of a PACE admission varies by the ranking within each high school based on the GPA in grades 9 and 10 (baseline GPA ranking). The circles are averages within 20 equally-sized bins of 5 percentiles, while the solid and dashed lines represent the predicted values and 95% confidence intervals using standard errors clustered at high school level.

whether they ever repeated a grade, residence in Santiago, school rurality, and type of high school track (academic or vocational).[15]

4.2 Regression Discontinuity Design

The Regression Discontinuity Design (RDD) exploits the within-school cutoff for a PACE admission offer to identify causal effects. The first criterion for a PACE admission causes a discontinuity in the probability of being offered a PACE admission, but this probability does not increase from zero to one due to the other admission criteria (see the description of the admission criteria in section 2.2). Therefore, we implement a Fuzzy Regression Discontinuity Design (Lee and Lemieux, 2010). We estimate the equation:

$$Y_{is} = \gamma + \delta A_i + f(p_{is} - c_s) + \theta X_i + \nu_i \tag{2}$$

where Y_{is} is the outcome for student i who attended high school s, A_i is equal to 1 if student i is offered a PACE admission and to 0 otherwise, and X_i is a vector of student

[15]There are 14,414 students in the 218 experimental schools at the beginning of the eleventh grade. For 61 of them we do not have a measure of baseline GPA ranking because they were not enrolled in a Chilean high school in the ninth and tenth grades. The share of students in the top 15% is 16.98% because there are students with the same GPA average at baseline.

characteristics.[16] We instrument A_i using the following first-stage regression:

$$A_i = \zeta + \phi I(p_{is} - c_s \geq 0) + g(p_{is} - c_s) + \psi X_i + u_i \tag{3}$$

where p_{is} is the GPA score used to rank student (PRN), i.e., the running variable, and c_s is the within-school cutoff for a preferential admission, defined as the minimum between the 85^{th} percentile of the PRN score in school s and a national cutoff. The standard errors are clustered at the school level. The functions $f(p_{is} - c_s)$ and $g(p_{is} - c_s)$ are linear functions of the score (normalised by the cutoff), and they are allowed to be different on either side of the cutoff. $I(p_{is} - c_s \geq 0)$ is an indicator function equal to 1 if student i's score is on the right side of the cutoff, and 0 otherwise.

For each outcome, we restrict the estimation sample to students within the optimal bandwidth around the cutoff computed according to Calonico, Cattaneo, and Titiunik (2014). We report parametric estimates as well as robust estimates with bias-correction obtained following Calonico, Cattaneo, and Titiunik (2014) and Calonico et al. (2019). The parameter δ identifies the impact on Y_{is} of being offered a PACE admission, for the subset of compliers near the cutoffs.

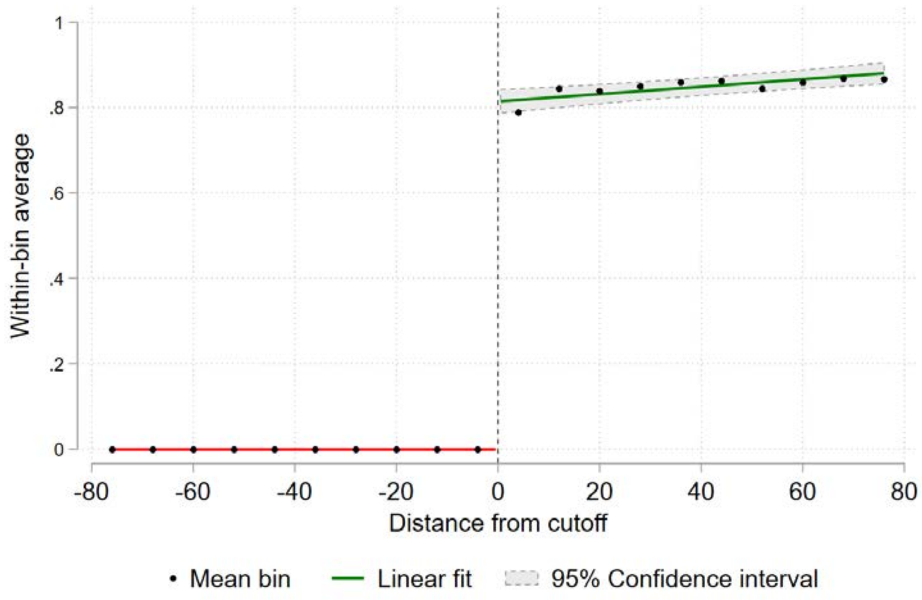
While no student below the cutoff is offered a PACE admission, not all students above the cutoff are offered one. Figure 4 displays the probability of receiving a PACE admission offer as a function of the distance of the PNR score from the cutoff (the 'First Stage'). No student with a score below the cutoff received a PACE admission, whereas 81% of the students above the cutoff received one, the probability slightly increases with the student's score.

Identification of the treatment effect requires the absence of manipulation in the ranking score. This assumption would fail if, for example, high school teachers adjusted their grading scheme in order to allow students' bunching above the national cutoff. To formally confirm the absence of ranking-score manipulation, we use the density of students as the dependent variable in Equation (3) (McCrary, 2008). The test does not reject the null hypothesis of no discontinuity in the density of students with a t-statistic of -0.390 and p-value of 0.697 (see Appendix Figure A4). We also find no evidence of discontinuity (p-value 0.940) when performing the manipulation test proposed by Cattaneo, Jansson, and Ma (2022), which does not require prebinning the data.

Identification of the treatment also requires comparable students on the two sides of the cutoff within the bandwidth (Lee and Lemieux, 2010). Table A4 reports the results of the t-test performed on the coefficient ϕ in Equation (3), using as a dependent variable one of the students' baseline characteristics. This table shows that only 1 out of the

[16] We use the same characteristics as those used in the RCT design. In the Appendix we show results without student characteristics as controls.

Figure 4: Probability of a PACE admission as a function of the distance of the PRN score from the preferential admission cutoff.



Notes: This figure shows how the likelihood of a PACE admission varies with the distance between a student's GPA score (PRN, defined in section 2.2) and the cutoff for a preferential PACE admission in her school. The circles are averages within 10 equally-sized bins on either side of the cutoff, while the solid and dashed lines represent the predicted values and 95% confidence intervals using standard errors clustered at high school level. The range of the score deviation corresponds to the optimal bandwidth computed following Calonico, Cattaneo, and Titiunik (2014).

11 estimations (age) is significant at conventional levels. Students above the cutoff are less than one month older than students below the cutoff. We do not believe that this represents a systematic difference between students marginally below and above the cutoff. The evidence suggests that the PACE admission offer is as good as (locally) randomly assigned around the cutoff.

5 The impacts of PACE on targeted students

We discuss the long-term impacts of PACE, separately for targeted students in this section and untargeted students in Section 6. Targeted students are those PACE aimed to bring to college: the top 15% of their school, including those around the preferential admission cutoff. We analyze the effect on these students in two ways: exploiting the RCT design by comparing the top 15% students in treated and control schools, and exploiting the RDD design by examining the students around the RDD cutoff.

5.1 Conceptual framework

It has long been posited that preferential admissions can have beneficial 'quality' effects and detrimental 'match effects' on the students they target (Arcidiacono and Lovenheim, 2016). Selective colleges provide larger monetary inputs, higher-quality professors and more academically skilled peers than non-selective institutions. But they are also more academically demanding: the courses may be more advanced and taught at a faster pace, professors may take more concepts and skills for granted, and the exams may be harder to pass. The so-called mismatch hypothesis (Sander, 2004) posits that disadvantaged students with low levels of academic preparation may reap larger long-term benefits from enrolling in non-selective institution than from enrolling in selective colleges. Affirmative action targeted at low-achieving students, according to this hypothesis, could make targeted students worse off by inducing them to enroll in selective colleges only to later drop out instead of graduating from non-selective institutions, lowering their lifelong earnings.

Mismatch could also occur among students who would enter the labor market right after high school absent preferential admissions. If preferential admissions induce them to enroll into higher education–an extensive margin effect–and later drop out, the short-term earning losses may not be offset by higher future wages. This kind of mismatch has received less attention in the literature, but could be particularly relevant in contexts, like PACE, that target students substantially disadvantaged and down the academic ability distribution, who would not normally enter higher education.

Existing evidence so far has mostly disputed this hypothesis, showing that preferential admissions to selective colleges have lasting benefits for disadvantaged students (Black, Denning, and Rothstein, 2023; Bleemer, 2022, 2021). PACE targets considerably less prepared students compared to the average admitted to selective colleges, reaching a population of students with lower pre-college achievement than those targeted by similar programs in other countries.[17] Whether mismatch can occur in such a population remains an empirical question. To answer it, this section analyzes the long-term impacts of PACE on the labor market outcomes of targeted students and explores mediating factors through the role played by education and individuals' expectations.

5.2 Students in the top 15% of baseline GPA

Labor market outcomes: modest positive impacts on long-term earning. Panel A of Figure 5 reports the impact of being in a school randomly selected to participate

[17]For example, students around the admission cutoff in the Eligibility in the Local Context percent plan in California obtain entrance exam scores that are above the average score among all college applicants (Bleemer, 2021), and the students induced to enroll in college by the Texas Top Ten percent plan score at the 89^{th} statewide percentile of the entrance exam distribution (Black, Denning, and Rothstein, 2023) and perform better academically than the untargeted students they displace (Kapor, 2024).

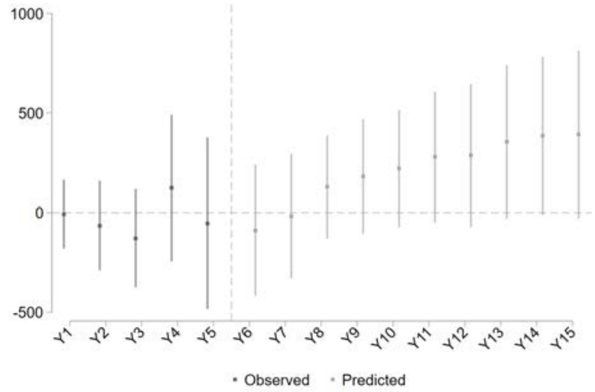
in PACE on students in the top 15% of their school based on GPA at the experiment's baseline. The left graph focuses on yearly earnings, reporting the estimate of β in equation (1) in each year after high school graduation. We observe labor market outcomes for students up to five years after high school graduation and we predict them up to year fifteen, using the surrogate index as explained in Appendix C.2. The right graph in Figure 5 presents the average yearly earnings of high achievers in treated and control schools in the short term (1-5 years after high school graduation), medium term (6-10 years after high school graduation), and long term (11-15 years after high school graduation). Table A5 reports the coefficients estimated in Panel A, while Table A6 reports the impact on labor force participation.[18]

During the first ten years after high school graduation, we find that on average PACE had no impact on the labor market outcomes of students in the top 15% of their school GPA, neither affecting the labor force participation nor the earnings, with some small and statistically insignificant impacts. However, despite no effect on months worked, we predict a positive impact on the earnings of high achievers in the long term: from 11 to 15 years after graduation, the yearly earnings of treated students are expected to be approximately 4% higher compared to control students, although these effects are statistically significant only at 10 percent level.

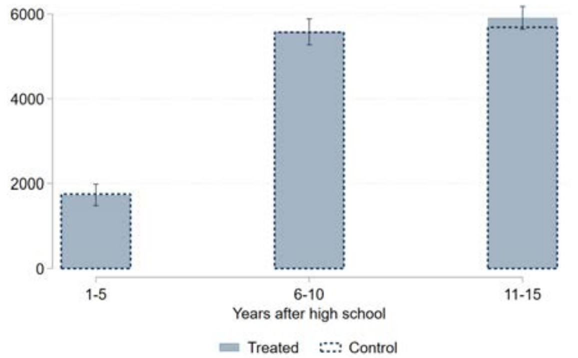
[18] Table A7 shows whether individuals appear either in the education or private sector data or both in the first five years after high school graduation. Overall, we observe more than 80% of students in higher education or private employment or both. There are only small differences between the treatment and control groups in year 1 and 5, statistically significant at 10 percent. Our long-term impact is not affected by data availability issues as the predictions using the surrogate index are calculated for all students in the dataset.

Figure 5: Labor market effect for top 15% – experimental sample

Panel A: All sample

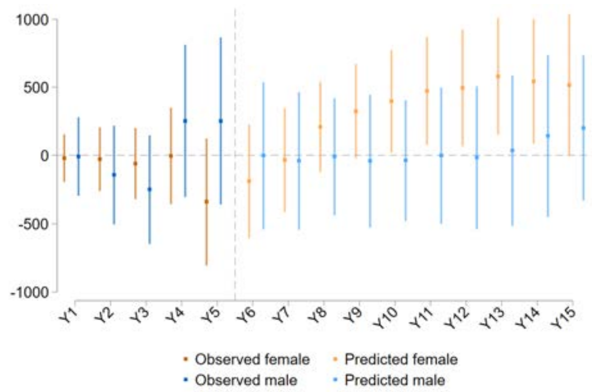


(a) Earnings, in a given year

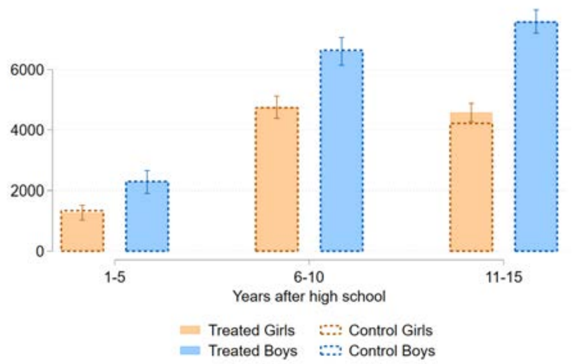


(b) Earnings 5-year average

Panel B: By gender



(c) Earnings, in a given year



(d) Earnings, 5-year average

Notes: This figure is based on data from the experimental sample of students who belong to the top 15% of GPA in their school at baseline (i.e. according to the GPA in grades 9 and 10). The left-hand-side graphs plot estimates of parameter β in equation (1) over time, while the right-hand-side graphs represent β as the difference between the outcome mean in the treatment and control groups. The x-axis indicates the year since high school graduation. The outcome variables are observed in years 1-5, and based on the surrogate index in years 6-15 (Appendix C.2). The bars represent 95% confidence intervals for β calculated from standard errors clustered at school level. For the outcomes based on the surrogate index, the standard errors are bootstrapped using 100 replications and resampling at school level. In the left-hand-side graphs, the dependent variable is earnings in a given year. In the right-hand-side graphs, the dependent variable is the average yearly earnings over the periods indicated on the x-axis. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). In Table A5, Panel A shows the estimated coefficients for Graph (a), Panel B for Graph (b). In Table A8, Panel A shows the estimated coefficients for Graph (c), Panel B for Graph (d).

Higher education outcomes: increased graduations from selective colleges. Figure A5a present the impacts on enrollment and graduation from higher education up to six years after the end of high school and Table 3 shows our main education outcomes including enrollment, graduation, and dropout from selective colleges (columns 1-3) and any tertiary institution (columns 4-6).

PACE anticipated entry into higher education for the top 15% sample rather than inducing new entry. While PACE increased enrollment in higher education substantially the first year after high school (Figure A5a , +6.2 p.p., a 10% increase), it did not increase the likelihood of ever enrolling during the six years post high school, nor of graduating or dropping out on average. This is consistent with the null impacts on labor force participation in the first five years after high school graduation and reflects the requirement that a PACE seat must be used the first year after high school, potentially substituting a gap year with earlier college enrollment.[19]

Table 3: ATE on education outcomes by higher-education institution *with controls* (RCT, baseline top 15%)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Ever enrolled	Graduation	Dropout	Ever enrolled	Graduation	Dropout
	A. Main results					
Treatment	0.114***	0.062**	0.045***	0.010	0.005	0.005
	(0.032)	(0.027)	(0.015)	(0.015)	(0.020)	(0.018)
Total obs.	2437	2437	2437	2437	2437	2437
Mean	0.423	0.278	0.103	0.860	0.604	0.257
	B. Heterogeneity by gender					
Treatment – Girls	0.153***	0.106***	0.050***	0.032*	0.027	0.004
	(0.034)	(0.033)	(0.018)	(0.017)	(0.025)	(0.022)
Mean Girls	0.437	0.297	0.097	0.879	0.650	0.230
Total obs.	1369	1369	1369	1369	1369	1369
Treatment – Boys	0.067	0.009	0.039*	-0.015	-0.023	0.008
	(0.044)	(0.035)	(0.023)	(0.022)	(0.029)	(0.029)
Mean Boys	0.405	0.253	0.110	0.837	0.545	0.292
Total obs.	1068	1068	1068	1068	1068	1068
p-value Treat Girls=Boys	0.054	0.019	0.697	0.077	0.162	0.897

Notes: In this table, we report the estimate for coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. Panel A reports the main results on the entire sample. Panel B reports the results for boys and girls, separately. At the bottom of the panel, we include the p-value for the test for whether the treatment effect on girls is different compared to the treatment effect on boys. *Mean* is the average of the outcome variable in the control group for each specific sample. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** p<0.01, ** p<0.05, * p<0.10.

Despite the null short- and medium-term impacts on labor force participation, earnings, and graduation rates from any higher education institution, we predict long-term earning gains based on students' characteristics and their labor market and educational paths up to five years after high school. The impact of PACE on selective college attendance helps shed light on these results.

Panel A of Table 3 shows PACE increased enrollment in selective colleges by 11.4 percentage points, but it also increased dropout from the same institutions by 4.5 percentage points (column 1 and 3 of Panel A). Hence, PACE ended up raising the likelihood

[19]This requirement was extended to two years after high school for cohorts younger than those in this study. The new requirement aligns with the regular application process, where the score on the PSU exam is valid for two years.

of graduation from selective colleges by 6.2 percentage points (column 2 of Panel A).[20] Interestingly, PACE did not increase dropout from higher education overall, indicating that the same share of students dropped out in treated and control school from any tertiary institution. PACE, therefore, increased the share of students graduating from selective college, and this did not come at the cost of increased overall dropouts from higher education.[21]

Who benefits? Not all students whose enrollment choices are affected by PACE ultimately obtain a degree from a selective college. This raises the question of which students experience the greatest benefits in terms of college degrees and labor market outcomes. To answer it, we explore effect heterogeneity.

In Panel B of Figure 5, we report the earning results by gender, providing clear evidence of substantial gender gaps. Among students in the control schools, women earn on average between 55 and 71 cents for every dollar earned by men, close to the national average of 58 cents for every dollar (World Economic Forum, 2023). This is at least partly driven by the fact they work between 15% and 35% fewer months than men (Table A12). Two striking facts emerge when we investigate the treatment effect of PACE on earnings by gender. First, the intervention had no impact on labor market outcomes neither for women nor for men in the short and medium terms.[22] Second, the positive impact on the long-term earnings is fully driven by women: the treatment increases their yearly earnings in the long term by 9 percent, while the effect is negligible for men. This implies that thanks to the participation in PACE, in the long term 11% of the gender gap in earnings among top-performing students was closed.

The results on education outcomes by gender, reported in Panel B of Table 3, help explain the gender differences in labor market impacts. The effects of PACE on enrollment and graduation from any tertiary institution are similar by gender (columns 4 and 5, see also Figures A6a and A6b), with qualitatively slightly positive results for women and negative for men. However, the magnitude is small and mostly insignificant at conventional levels, leading to similar labor market results in the short and medium terms. In

[20] The treatment effect on ever enrolled in selective college is mechanically equal to the sum of the treatment effects on enroll and graduate (column 2), enroll and drop out (column 3), and enroll and transfer to an institution of a different kind. The latter effect is not reported in the tables as it is mostly negligible. Regarding enrollment into any kind of institution, the treatment effect on ever enrolled (column 4) is mechanically equal to the sum of the treatment effects on enroll and graduate (column 5) and enroll and drop out (column 6), aside for rounding errors. Transfers are zero by definition in this case as we consider the higher education sector as a whole. Notice that in the RDD analysis, the effects across columns do not similarly add up because the estimation samples vary across columns due to different optimal bandwidths.

[21] All results presented in this section are robust even when we estimate them without control variables. As shown in Tables A9, A10 and A11, the results are qualitatively and quantitatively similar.

[22] There are no gender differences in the effect of PACE on the likelihood of appearing in the dataset in the first five years with some modest differences in selected years (Table A13).

contrast, the positive effect on selective college graduations is entirely driven by women, who increase their graduation from selective colleges by 10.6 percentage points, a 36% increase compared to the control group.[23] Consequently, based on the enhanced average performance in selective education for women, we predict average long-term earning gains for this group.

The results broadly relate to the literature on gender differences in preferential admission, but with stronger heterogeneity. For example, Black, Denning, and Rothstein (2023) investigating the impact of the Texas Top Ten find that BA degrees increase for women induced to attend by 6 percentage points, and only 2 percentage points for men, with a statistically different effect. Qualitatively, 9-11 years after high school graduation the improvement in earnings is almost three times as high for women compared to men, but the difference disappears in the long term in their context. Students in our context are more disadvantaged compared to those affected by the Texas Top Ten, and family disadvantage has been shown to be especially harmful for boys (Autor et al., 2019; Chetty et al., 2016; Carlana, La Ferrara, and Pinotti, 2022). Our evidence provides new insights that extend beyond childhood and adolescence: even when disadvantaged males become academic top-performers by the end of high school, they are less likely than their female counterparts to benefit from additional opportunities to attend college and to potentially improve their long-term social mobility.

5.3 Marginally eligible students: RDD results

Labor market outcomes: short-term earning losses, long-term earning gains. Next, we investigate the effects on lower-ability students around the cutoff for the assignment of PACE slots. The impact on them may differ compared to the average treatment for all high-achievers, influenced by whether 'quality' or 'match' effects are stronger.

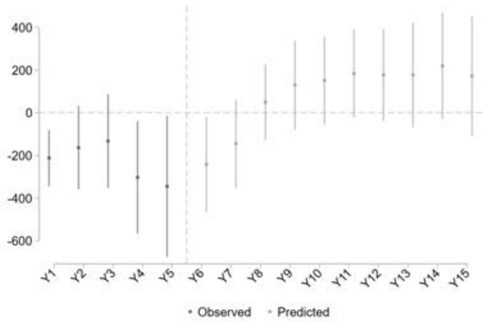
Figure 6 presents the labor market effects from the estimation of the second stage RDD (equation 2). Each bar in the left graphs reports the estimate of parameter δ in a given year in the RDD sample of students around the cutoff of top 15% of the PRN score (based on GPA over the four years of high school), while the right graphs reports the average yearly earnings in the short, medium, and long term. Tables A17 and A18 report the RDD estimate of being in the dataset and the months worked in the first five year after high school.[24]

[23] As selective colleges typically admit students with higher test scores compared to non-selective institutions, PACE heightened the selectivity of the programs in which the top-15% students enroll, as measured by the peers' average high school test scores (Table A14). The effect is stronger for women (Table A15) compared to men (Table A16).

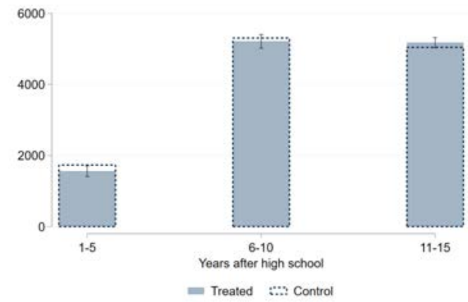
[24] We find no effects of appearing in the data within the first five years and significant effects only in the first two years.

Figure 6: Labor market effect – RDD sample

Panel A: All sample

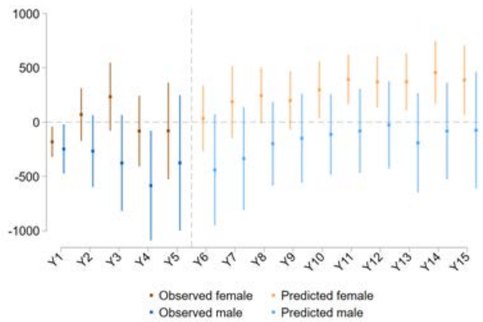


(a) Earnings, in a given year

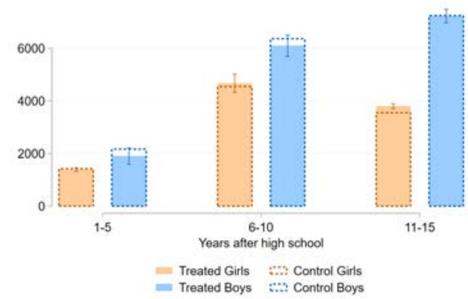


(b) Earnings, 5-year average

Panel B: By gender

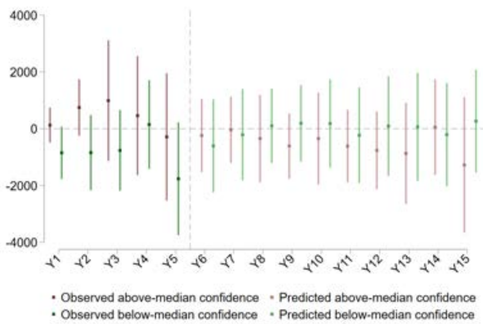


(c) Earnings, in a given year

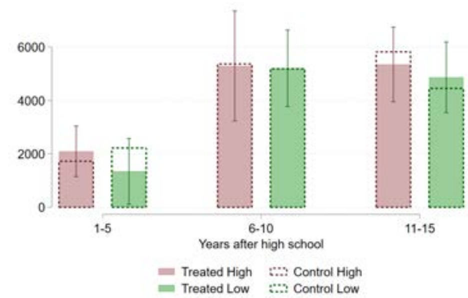


(d) Earnings, 5-year average

Panel C: By overconfidence



(e) Earnings, in a given year



(f) Earnings, 5-year average

Notes: This figure is based on data from the RDD sample of students around the preferential-admission cutoff in terms of the GPA score (PRN, defined in section 2.2). The left-hand-side graphs plot estimates of parameter δ in equation (2) over time, while the right-hand-side graphs represent δ as the difference between the outcome mean just below ("control") and just above ("treated") the cutoff for a preferential admission. The x-axis indicates the year since high school graduation. The outcome variables are observed in years 1-5, and based on the surrogate index in years 6-15 (Appendix C.2). The bars represent 95% confidence intervals for δ calculated from standard errors clustered at school level. For the outcomes based on the surrogate index, the standard errors are bootstrapped using 100 replications and resampling at school level. In the left-hand-side graphs, the dependent variable is earnings in a given year. In the right-hand-side graphs, the dependent variable is the average yearly earnings over the periods indicated on the x-axis. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). In Table A19, Panel A shows the estimated coefficients for Graph (a), Panel B for Graph (b). In Table A20, Panel A shows the estimated coefficients for Graph (c), Panel B for Graph (d). In Table A21, Panel A shows the estimated coefficients for Graph (e), Panel B for Graph (f).

The results on employment and earnings are strikingly different compared to the average effect for students in the top 15% of their school (Tables A18 and A19). When eligible for a PACE slot, marginal students have higher probability of being observed in the education or labor market dataset the first two years after high school graduation (Table A17) due to strong effects on enrollment, fewer short-term months of employment (-10.3%) and lower earnings (-10.9%). In the long term, we observe economically and statistically negligible effects of preferential admission on labor force participation and a 2.6% insignificant increase in yearly earnings.

Higher education outcomes: higher graduation from selective colleges, but also higher enrollment in tertiary education followed by dropout. The results on education outcomes shed light on the pattern behind the labor market results. Even for marginal students, PACE achieved on average its objective of increasing enrollment and graduation from selective colleges as shown in columns 1 and 2 of Panel A, Table 4. More broadly, the PACE admissions also increased the likelihood of enrolling altogether in tertiary education (a 4.1 p.p, or 5 percent, increase within six years of high school – column 4, Panel A), albeit to a lower extent than the increase in first-year enrollments.[25] We interpret these results as evidence that preferential admissions anticipated entry into higher education among some marginal students, like for the RCT top-15% sample, but they also induced new entry, a potential reason for the negative effect on labor force participation and earnings in the short term in this group.

A key difference with the RCT top-15% analysis emerges. While both analyses show an increase in graduation from selective colleges (column 2 of Panel A of Table 4 compared to Table 3), the sample of marginal students also exhibits an increase in the probability of enrolling and dropping out from higher education altogether, by 6.5 percentage points, a 27.8% increase (column 6 of the top Panel). The effect on the rate of graduation from higher education is negative, although statistically insignificant (column 5 of the top Panel). This suggests that some treated students who enroll in higher education and subsequently drop out might otherwise have not enrolled at all, potentially incurring earning losses, while others might have enrolled in different types of higher education institutions and graduated, possibly facing different job opportunities.[26] The results indicate

[25]Enrollment in higher education increased substantially the first year after high school (14.5 p.p., a 25 percent increase, Figure A7).

[26]While both the RCT-top15% and RDD analyses reveal a widening of the gap between students' test scores and the scores of the peers in the higher education programs they attend, as well as a worsening of student's test-score ranks within these programs, suggesting increased misalignment between program demands and students' preparation, these effects are more pronounced for marginal students compared to the average targeted student (Tables A14 and A27).

that preferential admissions may have had unintended consequences for some marginal students.[27]

Who benefits? Preferential admissions enhance the college attainment of some marginal students, and increase the likelihood of enrolling in higher education only to later drop out for others. This raises the questions of which students experience the benefits in terms of college degrees and which instead experience the increase in dropout, and how such diverse education impacts translate into labor market impacts. To answer them, we explore effect heterogeneity.

Men and women similarly reduce the months worked (Table A25) and increase their probability of being observed in the education or labor market dataset (Table A26) only in the first year after high school graduation, when offered a preferential admission. In Panel B of Figure 6 we report the earning results by gender. Graph (d) shows that the negative impact is driven by men during the first five years post high school, with a decrease of 12.3% in the average yearly earnings. Despite the negative impact on labor force participation, the effect for women's earnings is almost zero in the same period suggesting some potential positive effects on their wages. In the long term the earning effect for men is null, but for women we observe a pattern similar to the RCT top-15% sample, with a significant increase by 7.1% in yearly earnings (Table A20). The impacts, therefore, display striking gender differences: on average, men incur short-term earning losses while women incur long-term earning gains.

The results on education outcomes by gender, reported in Panel B of Table 4, help explain such gender differences in labor market impacts. Male and female students experience a similar increase in the likelihood of ever attending higher education, consistent with their reductions in employment in the short term. Both of them increase their likelihood to enroll and graduate (+0.054 p.p.) and enroll and drop out (+0.087 for women and +0.129 for men) from selective colleges.[28] But a key distinction is that the increase in dropouts from any tertiary institutions is entirely concentrated among males, who experience a 12.7 p.p. increase (almost a 50% increase), compared to a smaller and insignificant effect for females (column 6). Because the impact on enrolling to later drop out (column 6) is stronger than the impact on enrolling (column 4), the dropouts likely occurred among

[27] All results presented in this section are robust even when we estimate them without control variables. As shown in Tables A22, A23 and A24, the results are qualitatively and quantitatively similar.

[28] As selective colleges typically admit students with higher test scores than non-selective institutions, by increasing the likelihood of attending selective colleges preferential admissions heightened the selectivity of the programs attended among the group of always takers (who attend higher education regardless of preferential admission). Preferential admissions increased the selectivity of attended programs measured as the average high school test score of program peers, and the distance between a student's own test score and peers' test scores. They also decreased a student's own test score rank in the higher education program (Table A27). These effects are similar across genders (Tables A28 and A29).

Table 4: LATE on education outcomes by higher-education institution *with controls* (RD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Main results					
	Ever enrolled	Graduation	Dropout	Ever enrolled	Graduation	Dropout
Conventional	0.171***	0.055***	0.096***	0.041***	-0.023	0.065***
	(0.024)	(0.021)	(0.014)	(0.014)	(0.023)	(0.021)
Robust	0.163***	0.048**	0.096***	0.044***	-0.031	0.072***
	(0.026)	(0.023)	(0.017)	(0.017)	(0.026)	(0.023)
Bandwidth	53	64	75	69	53	56
Bandwidth obs.	10645	12682	14810	13668	10645	11233
Mean	0.402	0.259	0.090	0.857	0.618	0.234
	Heterogeneity by gender					
	Girls					
Conventional	0.155***	0.054**	0.087***	0.049***	0.011	0.034
	(0.034)	(0.027)	(0.018)	(0.018)	(0.028)	(0.022)
Robust	0.142***	0.045	0.083***	0.051**	0.006	0.035
	(0.038)	(0.030)	(0.020)	(0.021)	(0.033)	(0.026)
Bandwidth	39	58	73	61	58	79
Bandwidth obs.	4793	6910	8611	7196	6910	9225
Mean	0.410	0.269	0.076	0.867	0.652	0.222
	Boys					
Conventional	0.188***	0.054**	0.129***	0.044*	-0.042	0.127***
	(0.030)	(0.027)	(0.026)	(0.023)	(0.034)	(0.038)
Robust	0.179***	0.052*	0.136***	0.047*	-0.054	0.142***
	(0.033)	(0.031)	(0.029)	(0.027)	(0.038)	(0.042)
Bandwidth	78	83	61	75	60	47
Bandwidth obs.	6332	6710	4972	6095	4886	3864
Mean	0.386	0.240	0.108	0.834	0.565	0.258
p-value Girls=Boys	0.395	0.996	0.122	0.858	0.205	0.020
	Heterogeneity by overconfidence					
	High Overconfidence					
Conventional	0.162	-0.135	0.253***	0.008	-0.239**	0.348***
	(0.113)	(0.103)	(0.087)	(0.067)	(0.111)	(0.100)
Robust	0.179	-0.178	0.272***	0.015	-0.270**	0.369***
	(0.132)	(0.112)	(0.099)	(0.078)	(0.128)	(0.118)
Bandwidth	62	40	65	61	49	47
Bandwidth obs.	352	231	367	352	282	275
Mean	0.499	0.380	0.109	0.922	0.752	0.116
	Low Overconfidence					
Conventional	0.118	0.186	-0.046	-0.030	0.128	-0.179
	(0.137)	(0.120)	(0.117)	(0.092)	(0.155)	(0.129)
Robust	0.142	0.203	-0.009	-0.040	0.150	-0.190
	(0.162)	(0.137)	(0.135)	(0.107)	(0.179)	(0.150)
Bandwidth	78	77	47	95	68	72
Bandwidth obs.	292	290	186	363	256	272
Mean	0.340	0.169	0.128	0.781	0.465	0.350
p-value High=Low	0.742	0.029	0.032	0.749	0.047	0.002

Notes: In this table we report the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Mean* is the mean of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** p<0.01, ** p<0.05, * p<0.10.

both newly enrolled men and men who, absent the preferential admission, would have enrolled and graduated from vocational and non-selective programs, where their ability would have been more closely aligned to the program's academic demands. Therefore, among the men induced to enroll and drop out from higher education by a preferential admission, some likely experienced a delayed entry into the labor market, while others likely forewent job opportunities mostly reserved for graduates of higher education they would have otherwise had.

These results help explain the labor market effects for men. In the short and medium term, a higher share of men attend tertiary education losing their potential earnings, which are substantially higher than those of women. In the long term, two forces are at play leading to insignificant impact on earning: some men have positive impacts induced by higher graduation from selective colleges, while other men have lower earnings due to delayed entry in the labor market or foregone better-paying job opportunities. We interpret these results as evidence that preferential admissions induced a mismatch among some marginal men by increasing their likelihood of enrolling in higher education only to later drop out, causing short-term earning losses that may not be compensated by future higher wages.

Women too increase their dropout from selective colleges, but, as said, the overall dropout from any tertiary institution remained unchanged (columns 3 and 6, second Panel), suggesting women who dropped out from selective colleges would have dropped out from other institutions anyway. These results help explain the positive long-term impacts on earnings on average in this group, induced by higher graduation rates from selective colleges not accompanied by negative earning consequences from increased dropouts from higher education. Overall, the pattern for marginal women is similar compared to the sample of high-achievers, suggesting that when offered opportunities for higher income mobility through education, women are more likely to take full advantage of the opportunities offered to them.

Information frictions. Why would some marginal students, especially men, accept preferential admission offers from programs they will drop out from, potentially lowering their earnings in the long term? Information frictions might preclude them from foreseeing these negative consequences. This is a central tenet of the 'mismatch hypothesis' (Arcidiacono et al., 2011), that has so far been difficult to test directly absent expectation data. We test this hypothesis using expectation data we collected from students in the RCT sample at endline during their final high school year, and which we linked to their administrative records, including information on their outcomes up to six years later.[29]

[29] Table A30 shows that the sample with and without overconfidence measure is well balanced on all our baseline observables.

Figure A8 plots the histogram of the answers to a survey question that elicited students expected chances of graduating from a selective college if they were to enroll in one.[30] Half of the students in the sample are certain that enrolling in a selective college will guarantee their graduation, and this fraction exceeds 60 percent among those who eventually enroll in a selective college.

Contrasting high school students' perceived and actual chances of graduating from a selective college reveals a marked overoptimism regarding future college performance. Six years after the survey collection, a quarter of those who enrolled in a selective college had dropped out (Table A1), and this represents a lower bound on the dropout rate as 40% are still enrolled and could potentially still drop out.[31] Figure 7 displays a binscatter plot, comparing students' perceived graduation chances, based on numerical values assigned to the survey answer, to their actual chances of having graduated or being on track to graduate six years on, from the linked administrative data. Deviations from the 45-degree line indicate errors in beliefs. As can be seen, students along the distribution of true graduation chances are overoptimistic on average. However, Table A32 shows that men have on average almost 0.3 standard deviations higher overconfidence than women. The gender gap in overconfidence may be a key driver of differential effect of preferential admission on labor market and education by gender.

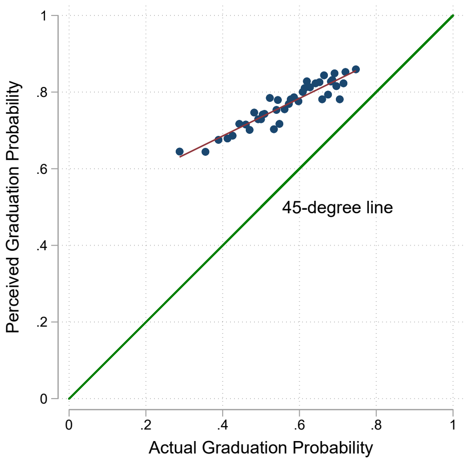
Despite the limited sample, we provide suggestive evidence on the mismatch hypothesis analyzing heterogeneity by overconfidence. If the negative impacts on the higher education performance of marginal students are driven by information frictions, they should be concentrated among overconfident students. Panel C of Table 4 shows that the preferential admission offer increased the likelihood of enrolling only to later drop out among those with above median overconfidence. These students qualitatively are more likely to have an earning premium in the short term and suffer a penalty in the long term compared to equally overconfident students in the control group (Panel C, Figure 6). In contrast, students who held more realistic expectations faced positive effects on their education outcomes when offered additional opportunities to attend college, although not significant at conventional level, including in their probability of graduation from selective colleges.[32] We observe results qualitatively consistent with this evidence on the labor market outcomes with a short-term earning loss for this group and long-term gains, likely

[30]See Appendix C.1 for the exact wording of the question.

[31]Table A31 further shows that nearly 40% of college entrants certain to graduate had instead abandoned college six years later.

[32]Preferential admissions also affected the distance between a student's baseline test score and the average baseline test scores of the peers in the higher education program. We find qualitatively stronger impacts on this 'ability distance' measure among more overconfident students, suggesting they were induced to enroll in programs for which they were relatively less prepared compared to less overconfident students (see Tables A33 and A34), consistent with the results on dropout. As selectivity is observed only conditional on enrolling in higher education, we present Lee (2009) bounds on the selectivity impacts for the sample of students who would enroll regardless of preferential admission.

Figure 7: Perceived and actual graduation chances.



Notes: This figure is based on students who were surveyed on their beliefs regarding selective college graduation. It presents a binscatter plot comparing students' perceived graduation chances (from a survey in their final high school year) with their actual chances of having graduated or being on track to graduate six years later (from linked administrative data). Students' perceived chances were assigned values of 0, 0.25, 0.50, 0.75, or 1 based on their survey responses. In Figure A9 we show robustness to the choice of numerical scale. The actual probabilities were predicted using a LASSO Probit model estimated on the sample of students who enrolled in selective colleges during the six years post high school. The data is grouped into 40 equal bins, each representing 2.5% of the total sample of 5,770 students with non-missing survey answers and LASSO-based predictions, and containing students with similar actual graduation probabilities. Each dot on the plot shows the average perceived graduation probability for the students in that bin. A 45-degree reference line is included; deviations from this line indicate discrepancies between students' beliefs and actual outcomes.

due to higher graduation from selective colleges.[33] This evidence is based on a limited number of observations, since the survey data was only collected for one cohort and only in schools that were part of the randomized expansion.[34] Nonetheless, it is consistent with the theory that affirmative action and preferential admissions can lead to mismatch effects in the presence of information frictions for low-achieving and for overoptimistic students (Arcidiacono et al., 2011). Mismatch, however, appears to be a local rather than a global phenomenon, even in a context in which the academic requirements for an admission are substantially relaxed.

6 The impact of PACE on untargeted students

Conceptual framework. Students belonging to the bottom 85% of the grade distribution in their school were not directly targeted. But preferential admissions can influence

[33]The results on labor market and education impacts by overconfidence are likely not driven by the assignment of numerical values to the Likert scale used to elicit subjective expectations. We obtain similar results when splitting the sample by whether students reported being certain that they will graduate from a selective college (approximately 50% of respondents), or whether they reported being less than certain. See Table A35 and Figure A10.

[34]Notice also that we cannot provide solid evidence using the RCT identification strategy as the awareness of treatment may have affected the overconfidence of students given that the measure was collected at the end of high school, two school-years after the start of the experiment.

classmates of targeted students through at least three channels that may act in different directions. First, exposure to peers that attend college may lead them to pursue tertiary education at higher rates induced by an imitation effect or desire to continue education like their friends (Golightly, 2019; Fernández, 2021; Anelli and Peri, 2019). Second, not being awarded a preferential admission may act as a negative signal on own academic ability and constrain aspirations, leading them to disengage from higher education and invest in labor market prospects (Genicot and Ray, 2017; La Ferrara, 2019). Finally, admission policies could generate equilibrium effects in local labor markets by increasing the supply of college entrants (Moretti, 2004). In this section, we analyze the causal effects of PACE on the long-term outcomes of bottom 85% students and investigate likely mediating factors.

Higher short-term labor force participation. For students belonging to the bottom 85% of the grade distribution in their school we find sizeable positive impacts on labor force participation and positive but insignificant impacts on earnings in the first five years after high school. Table A36 and Figure 8 presents the results on labor force participation and earnings, respectively.[35] In the first five years after high school graduation, students in the bottom 85% of GPA distribution in treated schools work on average 6.5% more months compared to students in the control schools every year. The effect is positive for all five years, but strong and statistically significant only in the first three years. The impact is relevant not only in terms of months worked, but also on the yearly earnings with a 7.6% higher amount compared to control students, even if more noisily estimated. The initial advantage in the labor market seem to persist over time, with positive point estimates on the labor force participation and predicted earnings up to fifteen years after high school graduation, although the effect is not statistically significant. Overall, students in treated compared to control schools are marginally better off considering their higher short-term earnings.

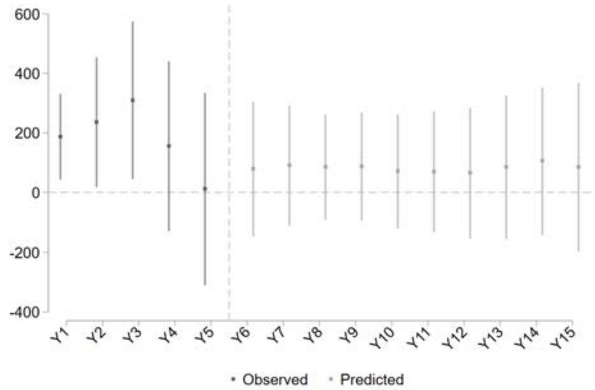
No positive education spillover effects. We find no impacts on the higher education outcomes for these students (see Table 5 and Figure A5b).[36] The results are stable over time and type of tertiary institution. The lack of increased enrollment in higher education suggests that PACE did not generate positive spillover effects on the educational achievement of untargeted students. Such effects were documented in other contexts, such

[35] Table A37 shows that there is no average impact of preferential admission on the probability of observing students in the education or private labor market in the short term (years 1-5). There is some small impact only in the first few years. Furthermore, Tables A38 and A39 provides evidence that the results are not qualitatively affected by the inclusion of controls.

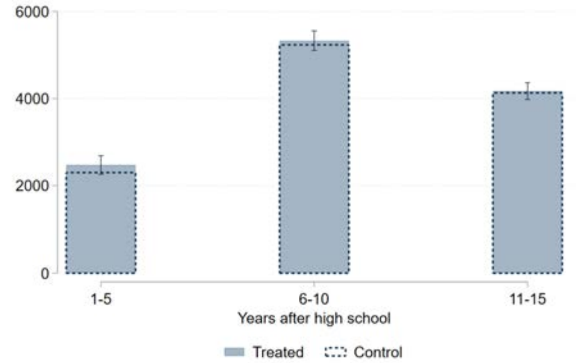
[36] These results are very similar regardless of the inclusion of control variables in the regressions (Table A40).

Figure 8: Labor market effect for bottom 85% – experimental sample

Panel A: All sample

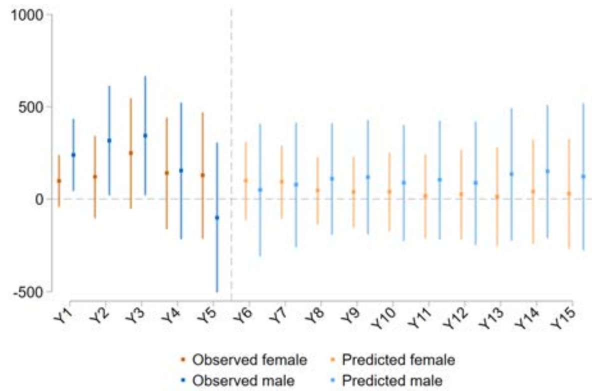


(a) Earnings, in a given year

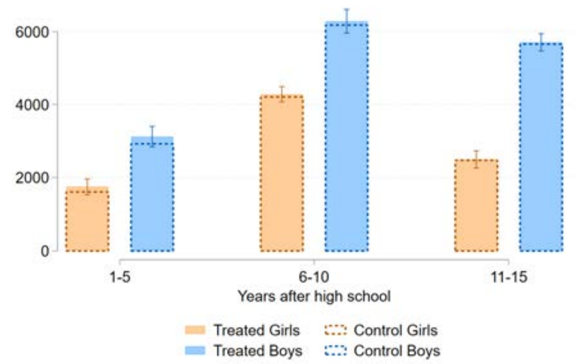


(b) Earnings, 5-year average

Panel B: By gender



(c) Earnings, in a given year



(d) Earnings, 5-year average

Notes: This figure is based on data from the experimental sample of students who belong to the bottom 85% of GPA in their school at baseline (i.e. according to the GPA in grades 9 and 10). The left-hand-side graphs plot estimates of parameter β in equation (1) over time, while the right-hand-side graphs represent β as the difference between the outcome mean in the treatment and control groups. The x-axis indicates the year since high school graduation. The outcome variables are observed in years 1-5, and based on the surrogate index in years 6-15 (Appendix C.2). The bars represent 95% confidence intervals for β calculated from standard errors clustered at school level. For the outcomes based on the surrogate index, the standard errors are bootstrapped using 100 replications and resampling at school level. In the left-hand-side graphs, the dependent variable is earnings in a given year. In the right-hand-side graphs, the dependent variable is the average yearly earnings over the periods indicated on the x-axis. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). In Table A43, Panel A shows the estimated coefficients for Graph (a), Panel B for Graph (b). In Table A44, Panel A shows the estimated coefficients for Graph (c), Panel B for Graph (d).

as that of the Texas percent plan (Golightly, 2019).[37] The result is likely to be driven by the wider gap in ability between the untargeted students and the regular entrants to university in the context of PACE compared to other programs, such as the Texas percent plan.

For completeness, we also report the heterogeneity by gender to investigate whether PACE led to differential spillovers. The impact of PACE on the bottom 85% of students by gender are reported in Table A41 and Panel B of Figure 8 for labor force participation and earnings, respectively. For all labor market outcomes, we do not detect statistically significant differences across genders.[38] Finally, consistently with the labor market results, we find no evidence of differential effects by gender among students in the bottom 85% of their high school GPA (Panel B of Table 5 and Figures A6c and A6d). These results suggest overall that PACE did not affect untargeted students deferentially by gender.

No evidence of ineligibility as a negative signal on academic ability. A student who is not offered a PACE admission slot discovers that she does not rank among the top-performing students in her school. This information can serve as a negative signal about own academic ability, potentially leading students in the bottom 85% of treated schools to revise downwards their beliefs about their academic skills. Consequently, these students may invest less in further education while increasing their focus on labor market opportunities. This is a plausible mechanism because of the absence of relative-rank feedback in these schools. Additionally, Tincani, Kosse, and Miglino (2024) documented that students in this setting are substantially overoptimistic about their within-school rank, with 43% of students in PACE schools that believe they are in the top 15% of the within-school GPA distribution. Therefore, receiving explicit notification of not being in the top 15% of the grade distribution is an informative signal that could alter students' perceptions and choices.

To study this channel, we examine whether students in treated schools reduce their engagement with the college application system in the years following the first one after high school, a clear signal that they are disengaging with higher education to potentially engage more in the labor market. Table A45 presents the results. We find precise zero effects, suggesting this behavioral mechanism is unlikely to be driving the positive impacts on these students' labor force participation right after high school graduation.

[37] See also Fernández (2021) for evidence of positive spillover effects of financial aid on the higher education of neighbours.

[38] Table A42 reports some slight differences in the probability of being observed in the dataset only for boys that may deserve further investigation.

Table 5: ATE on education outcomes by higher-education institution *with controls* (RCT, bottom 85%)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Ever enrolled	Graduation	Dropout	Ever enrolled	Graduation	Dropout
	A. Main results					
Treatment	0.021	0.007	0.009	-0.009	-0.003	-0.006
	(0.017)	(0.011)	(0.007)	(0.017)	(0.018)	(0.013)
Total obs.	11916	11916	11916	11916	11916	11916
Mean	0.138	0.085	0.039	0.672	0.390	0.282
	B. Heterogeneity by gender					
Treatment – Girls	0.022	0.009	0.010	-0.014	-0.022	0.007
	(0.020)	(0.016)	(0.007)	(0.020)	(0.023)	(0.014)
Mean Girls	0.158	0.109	0.034	0.745	0.497	0.248
Total obs.	5633	5633	5633	5633	5633	5633
Treatment – Boys	0.020	0.005	0.009	-0.004	0.013	-0.017
	(0.018)	(0.010)	(0.010)	(0.021)	(0.018)	(0.019)
Mean Boys	0.120	0.062	0.043	0.605	0.291	0.314
Total obs.	6283	6283	6283	6283	6283	6283
p-value Treat Girls=Boys	0.907	0.786	0.917	0.665	0.106	0.228

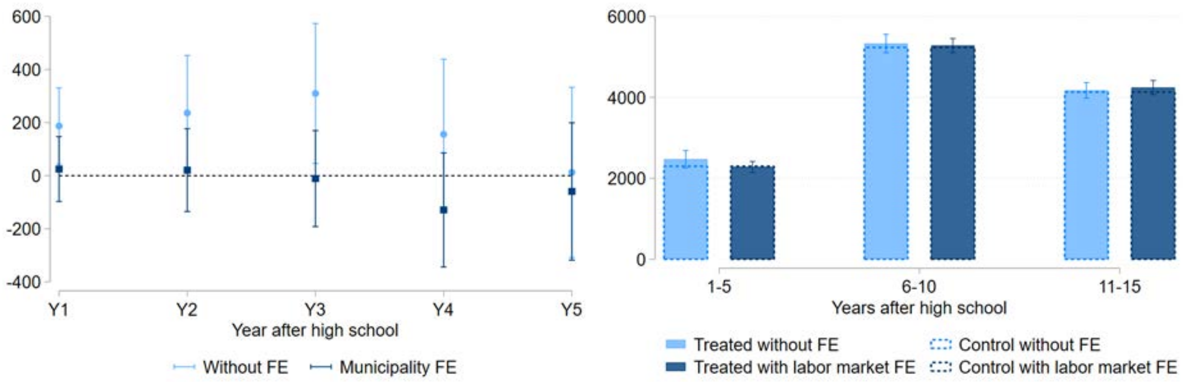
Notes: In this table, we report the estimate for coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. Panel A reports the main results on the entire sample. Panel B reports the results for boys and girls, separately. At the bottom of the panel, we include the p-value for the test for whether the treatment effect on girls is different compared to the treatment effect on boys. *Mean* is the average of the outcome variable in the control group for each specific sample. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Evidence of equilibrium effects in local labor markets. The first year after high school, top-15% students from treated schools are more likely to enter higher education than top-15% students from control schools, as shown in Figure A11. This difference potentially creates more job vacancies in the local labor markets where treated schools are located. In this initial phase, therefore, lower competition for jobs may increase the labor force participation of untargeted students in treated schools. From the third year forward, there may again be more vacancies available to untargeted students in local labor markets hosting treated schools, because top-performing students from treated schools attend longer degree programs and enter the labor market later than their counterparts in control schools. Equilibrium effects in local labor markets, therefore, are a possible mechanism behind the positive effects of PACE on untargeted students' labor force participation and earnings in the first few years after high school.

To study this channel, we examine whether such positive effects are robust to the inclusion of local labor market fixed effects. Under this channel, bottom-85% students from control schools that are located in the same local labor markets as at least one treated school face the same labor market conditions as their counterparts in treated schools. We should not observe PACE effects when comparing treated and control schools in the same local labor market. Therefore, evidence in favor of this channel would be that the treatment effects vanish with the inclusion of local labor market fixed effects.

Figures 9 and A12 provides evidence in support of the equilibrium effects in the local labor market. The positive impact on labor force participation and average yearly earnings observed in our standard regressions disappear when we take into account the municipality fixed effects. The result is robust to the choice of the fixed effect included to define the local labor market, as shown by Tables A46 for the average yearly months and by Table A47 for the average yearly earnings with different fixed effects at the municipality, province, and region level.

Figure 9: Earnings effects for bottom 85% with local labor market fixed effects — experimental sample



(a) Earnings, in a given year

(b) Earnings, average across years

Notes: This figure shows the estimates of parameter β in equation (1) over time, in the experimental sample of students who belong to the bottom 85% of their school at baseline (i.e. according to the GPA in grades 9 and 10). The x-axis indicates the year since high school graduation. The outcome variables are observed over the time period reported. The bars represent 95% confidence intervals for β calculated from standard errors clustered at school level. In Graph (a), the dependent variable is earnings in a given year. In Graph (b), the dependent variable is the average yearly earnings over the periods indicated on the x-axis. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). The regressions with fixed effects include fixed effects for the school's municipality. Panels A and B of Table A47 show the estimated coefficients for Graph (a) and Graph (b), respectively.

7 Conclusions

Preferential admission to college is potentially a powerful policy tool to enhance social mobility and compensate for inequality of opportunities earlier in life (Cunha and Heckman, 2007). In this paper, we investigate whether preferential admission could be extended to students with low academic preparation, while still delivering on its promises of improving long-term outcomes. We provide evidence from PACE (*Programa de Acompañamiento y Acceso Efectivo a la Educación Superior*), a top-percent plan implemented in Chile. First, we show that preferential college admissions can improve long-term earnings of targeted women by improving their graduation from selective colleges, without increasing overall dropout from any tertiary institution. We provide robust evidence of this key result using two different identification strategies: we compare high-achievers in schools randomly allocated to treatment and control as well as students around the discontinuity cutoff in schools that implemented the policy. Second, we show that preferential admission has a more nuanced effects on men that are generally more overconfident compared to women. While on average for the high-achievers there is no significant impact on labor market or education outcomes, among the marginal students men are more likely to enroll and dropout from tertiary institutions when extended a preferential admission, leading to short-term earnings losses, likely without offsetting long-term earning gains. There is a limit to how far inclusion can go while still delivering on its promises when students are not well-prepared for demanding tertiary education and are overconfident in their own abilities. Finally, we investigate the spillover effects on untargeted students, comparing the long-term outcomes for the bottom 85% of students in treated and control schools. We find positive impacts on their labor force participation, likely driven by general equilibrium effects in local labor markets.

The results from our study shift the attention in the research and policy debate from *whether* there is a mismatch due to affirmative action to *when* the mismatch occurs. They also illustrate that equilibrium spillovers on untargeted groups should be carefully considered, as they can increase the cost-effectiveness of preferential admission policies but also threaten the validity of difference-in-differences strategies commonly used to analyze them. Further research is necessary to investigate whether interventions aimed at adjusting the expectations or enhancing the academic preparation of targeted students can help deliver on the promises of improving social mobility through affirmative action.

References

Anelli, Massimo and Giovanni Peri. 2019. "The effects of high school peers' gender on college major, college performance and income." *The Economic Journal* 129 (618):553–602.

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

Arcidiacono, Peter, Esteban M Aucejo, Hanming Fang, and Kenneth I Spenner. 2011. "Does affirmative action lead to mismatch? A new test and evidence." *Quantitative Economics* 2 (3):303–333.

Arcidiacono, Peter and Michael Lovenheim. 2016. "Affirmative action and the quality-fit trade-off." *Journal of Economic Literature* 54 (1):3–51.

Athey, Susan, Raj Chetty, Guido W Imbens, and Hyunseung Kang. 2019. "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely." Tech. rep., National Bureau of Economic Research.

Autor, David, David Figlio, Krzysztof Karbownik, Jeffrey Roth, and Melanie Wasserman. 2019. "Family Disadvantage and the Gender Gap in Behavioral and Educational Outcomes." *American Economic Journal: Applied Economics* 11 (3):338–81. URL https://www.aeaweb.org/articles?id=10.1257/app.20170571.

Bailey, Martha J. and Susan Dynarski. 2011. "Inequality in Postsecondary Education." In *Whither Opportunity*, edited by Greg J. Duncan and Richard J. Murnane. New York: Russell Sage, 117–132.

Belley, Philippe and Lance Lochner. 2007. "The changing role of family income and ability in determining educational achievement." *Journal of Human capital* 1 (1):37–89.

Black, Sandra E, Jeffrey T Denning, and Jesse Rothstein. 2023. "Winners and losers? the effect of gaining and losing access to selective colleges on education and labor market outcomes." *American Economic Journal: Applied Economics* 15 (1):26–67.

Bleemer, Zachary. 2021. "Top percent policies and the return to postsecondary selectivity." *Research & Occasional Paper Series: CSHE* 1.

———. 2022. "Affirmative action, mismatch, and economic mobility after California's Proposition 209." *The Quarterly Journal of Economics* 137 (1):115–160.

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocío Titiunik. 2019. "Regression Discontinuity Designs Using Covariates." *The review of economics and statistics* 101 (3):442–451.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik. 2014. "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82 (6):2295–2326.

Carlana, Michela, Eliana La Ferrara, and Paolo Pinotti. 2022. "Goals and Gaps: Educational Careers of Immigrant Children." *Econometrica* 90 (1):1–29. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17458.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2022. "lpdensity : Local Polynomial Density Estimation and Inference." *Journal of statistical software* 101 (2):1–25.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income segregation and intergenerational mobility across colleges in the United States." *The Quarterly Journal of Economics* 135 (3):1567–1633.

Chetty, Raj, Nathaniel Hendren, Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi. 2016. "Childhood environment and gender gaps in adulthood." *American Economic Review* 106 (5):282–288.

Cunha, Flavio and James Heckman. 2007. "The technology of skill formation." *American economic review* 97 (2):31–47.

Dong, Yingying. 2019. "Regression Discontinuity Designs With Sample Selection." *Journal of business and economic statistics* 37 (1):171–186.

Dube, Arindrajit, Laura Giuliano, and Jonathan Leonard. 2019. "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior." *The American economic review* 109 (2):620–663.

Fernández, Andrés Barrios. 2021. "Neighbors' effects on university enrollment." *American Economic Journal: Applied Economics (forthcoming)* .

Genicot, Garance and Debraj Ray. 2017. "Aspirations and inequality." *Econometrica* 85 (2):489–519.

Golightly, Eleanor. 2019. "Does College Access Increase High School Effort? Evaluating the Impact of the Texas Top 10% Rule." *Mimeo, U of Texas Austin* .

Hanushek, Eric A, Jacob D Light, Paul E Peterson, Laura M Talpey, and Ludger Woessmann. 2022. "Long-run trends in the US SES—Achievement gap." *Education Finance and Policy* 17 (4):608–640.

Heckman, James J, John Eric Humphries, and Gregory Veramendi. 2018. "Returns to education: The causal effects of education on earnings, health, and smoking." *Journal of Political Economy* 126 (S1):S197–S246.

Horn, Catherine L. and Stella M. Flores. 2003. "Percent Plans in College Admissions: A Comparative Analysis of Three States' Experiences." The Civil Rights Project, Harvard University.

———. 2015. "Texas Top Ten Percent Plan: How It Works, What Are Its Limits, and Recommendations to Consider." Educational Testing Service.

Kapor, Adam. 2024. "Transparency and Percent Plans." Tech. rep.

La Ferrara, Eliana. 2019. "Presidential address: Aspirations, social norms, and development." *Journal of the European Economic Association* 17 (6):1687–1722.

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *Review of Economic Studies* 76 (3):1071–1102.

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

Machado, Cecilia, Germán Reyes, and Evan Riehl. 2023. "The direct and spillover effects of large-scale affirmative action at an elite Brazilian university." *forthcoming, Journal of Labor Economics* .

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

Ministry of Education. 2023. "Sistema de acceso y SUA desde 2016 100423." Tech. rep., Chilean Ministry of Education. Report of the Chilean Ministry of Education.

Moretti, Enrico. 2004. "Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data." *Journal of econometrics* 121 (1-2):175–212.

Neilson, Christopher, Federico Huneeus, Conrad Miller, Seth Zimmerman et al. 2021. "Firm Sorting, College Major, and the Gender Earnings Gap." Tech. rep.

Otero, Sebastián, Nano Barahona, and Cauê Dobbin. 2023. "Affirmative action in centralized college admission systems: Evidence from Brazil." *Unpublished manuscript* .

Sander, Richard H. 2004. "A Systemic Analysis of Affirmative Action in American Law Schools." *Stanford law review* 57 (2):367–483.

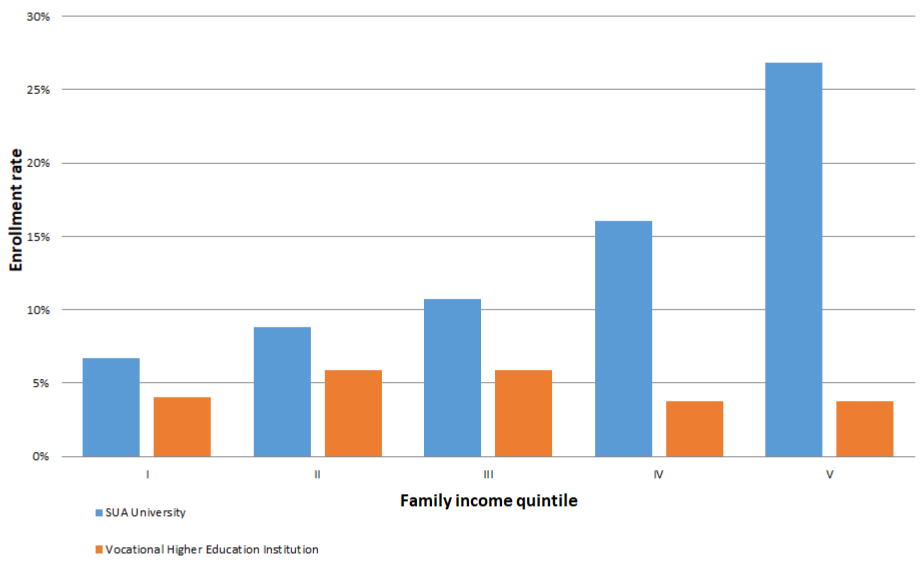
Tincani, Michela M, Fabian Kosse, and Enrico Miglino. 2024. "College Access When Preparedness Matters: New Evidence from Large Advantages in College Admissions." *mimeo, UCL* .

World Economic Forum. 2023 .

Appendix

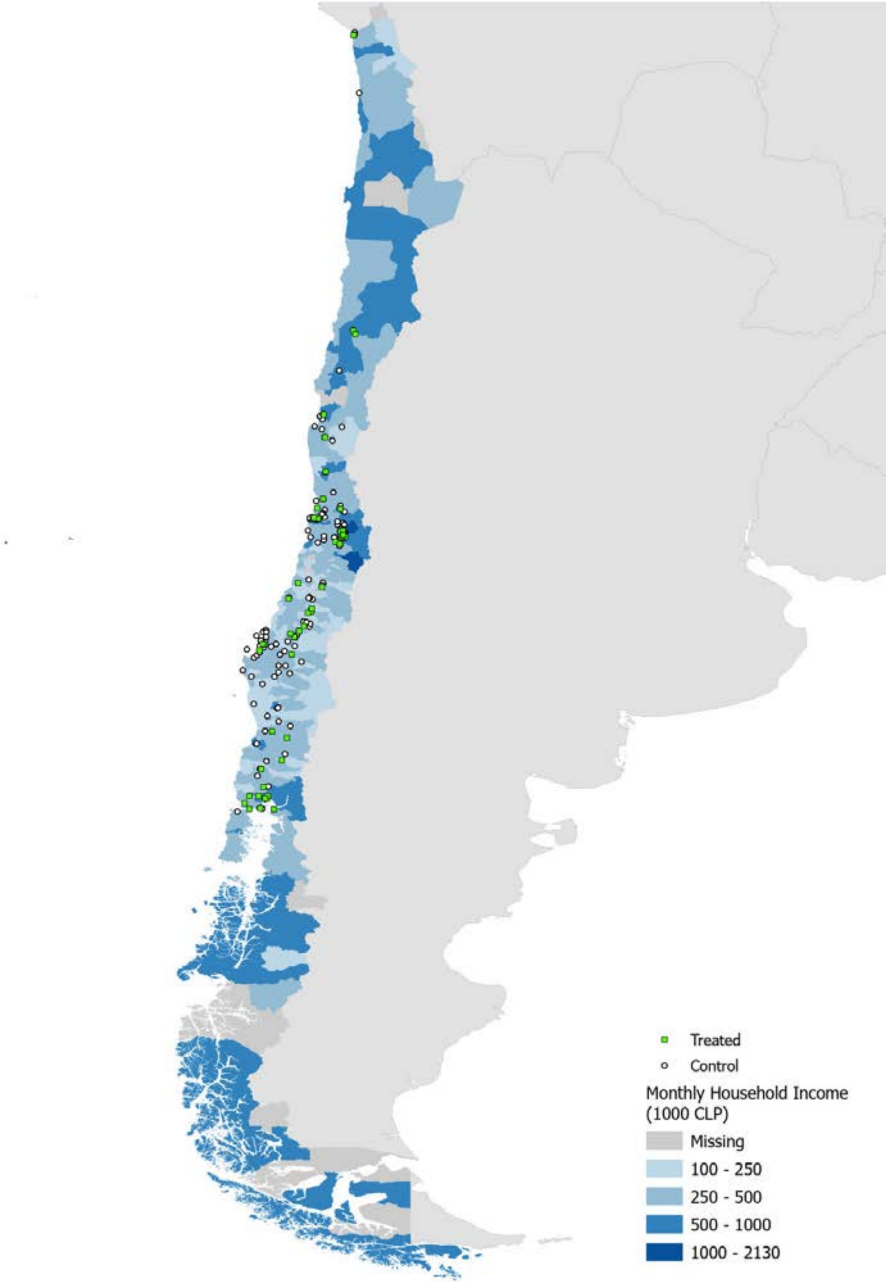
A Additional Figures

Figure A1: Socioeconomic inequality in selective college enrollment.



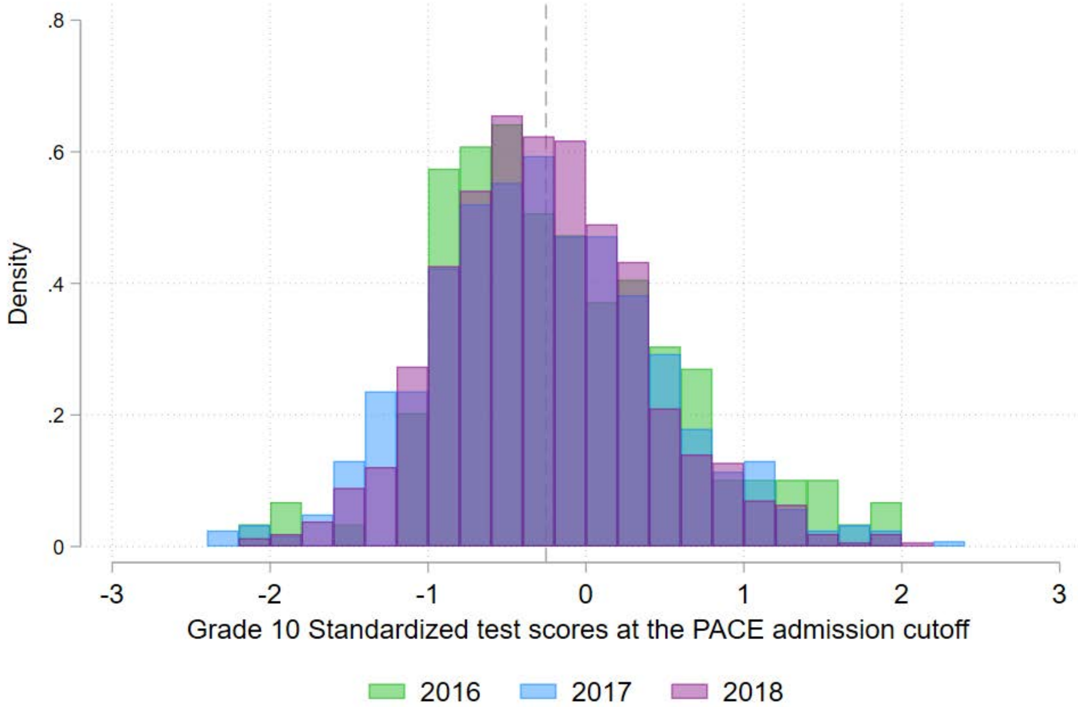
Notes: Percentage of 18-19 year old individuals who are enrolled in selective colleges and in vocational institutions in Chile, by family income quintile. Source: CASEN household survey, 2009, 2011 and 2013 waves.

Figure A2: Geographic distribution of schools in the experiment.



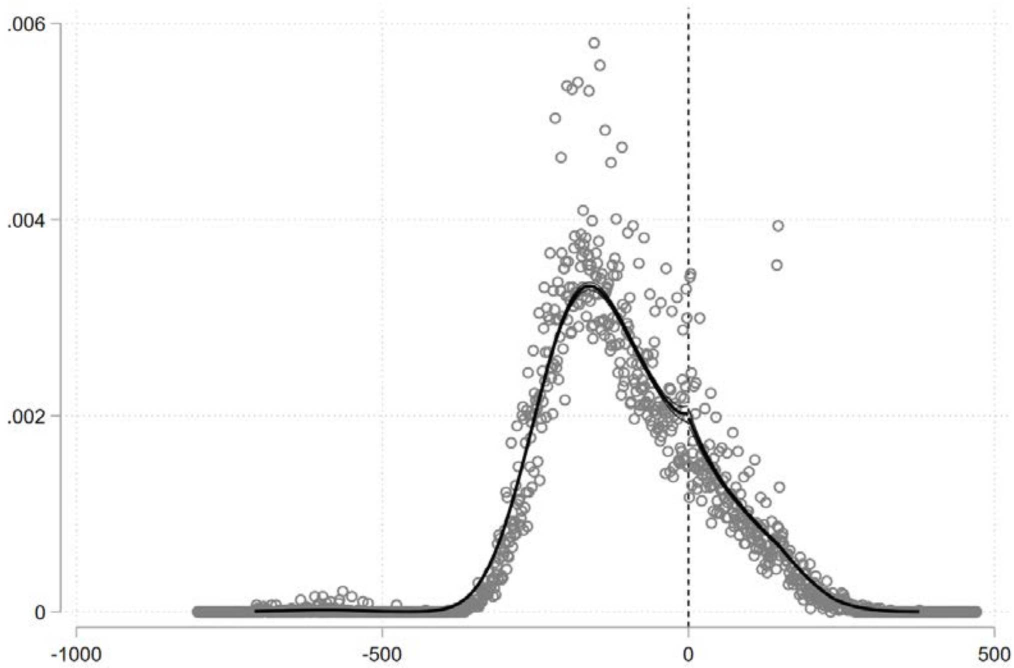
Notes: This figure shows the heatmap of Chile in terms of average monthly household income (1000 Chilean Pesos) and the geographic distribution of treated and control schools in the randomized experiment.

Figure A3: Academic preparation of PACE school students around the preferential admission cutoffs.



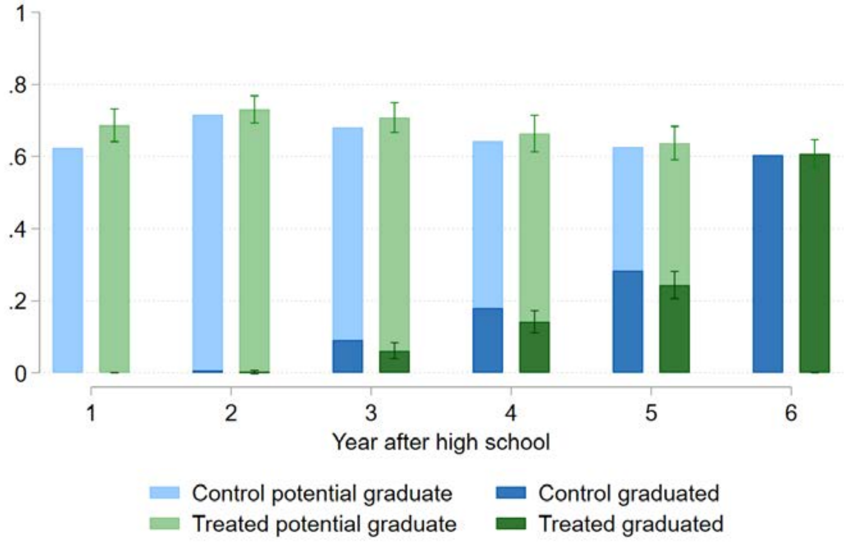
Notes: This figure shows the distribution of grade 10 SIMCE standardized test scores of students around the preferential admission cutoff. We obtained the SIMCE test score of the students whose school grades place them around the preferential admission cutoff in each school (within the optimal bandwidth in the first stage regression), and calculated their average test score. The figure plots the distribution of standardized scores at the cutoff in the three cohorts used in the RDD analysis, identified by the high school graduation year.

Figure A4: McCrary test.

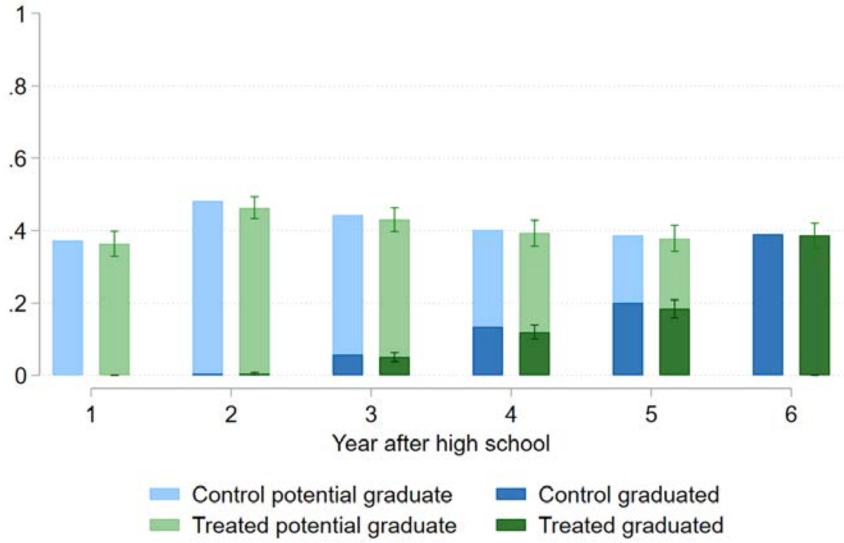


Notes: This figure shows the density of students in 1 score-point bins. The solid line plots fitted values from a local linear regression of density on ranking-score deviations from the cutoff, separately estimated on both sides of the cutoff. The thin lines represent the 95% confidence interval.

Figure A5: Effects of PACE on higher education enrollment and graduation over time, RCT analysis.



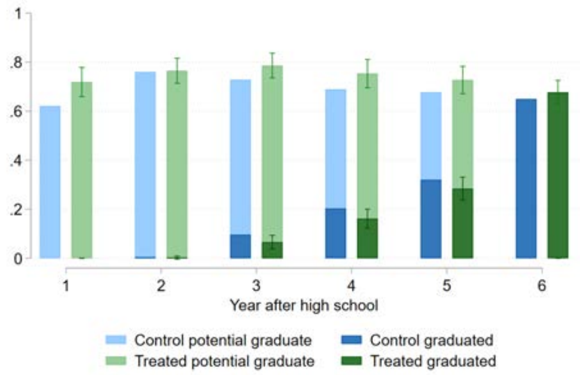
(a) Top 15% RCT sample



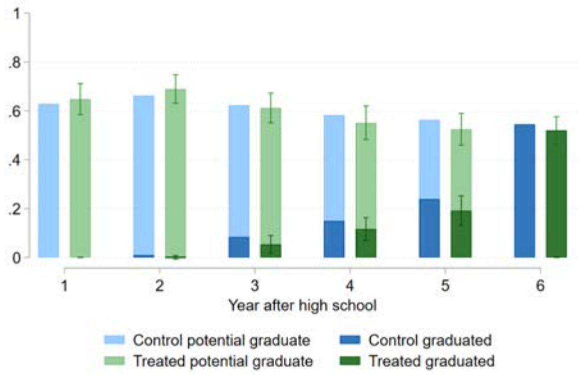
(b) Bottom 85% RCT sample

Notes: Define a potential graduate as a student who is enrolled or who has graduated in the current year or in a previous year. This figure plots, for every year after high school, the fraction of top 15% and bottom 85% students at baseline in control and treated schools that are potential graduates of higher education (light shade) or that have graduated from higher education (dark shade). The error bars represent 95% confidence intervals for the difference in proportions across treatment groups (i.e., the treatment effect). Graduation data in the sixth year is not yet available for the RCT sample; the sixth-year bars represent the fraction of students who are either still enrolled in the sixth year or who have graduated during the prior five years.

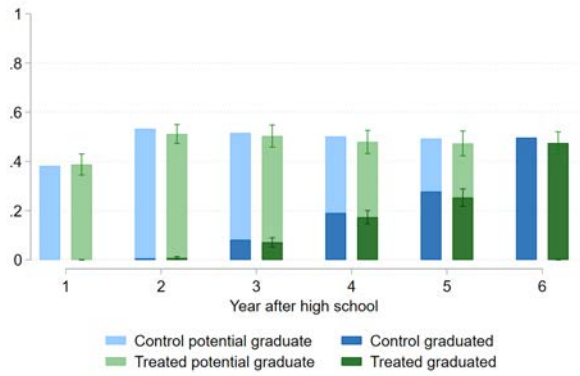
Figure A6: Effects of PACE on higher education enrollment and graduation over time, RCT analysis *by gender*.



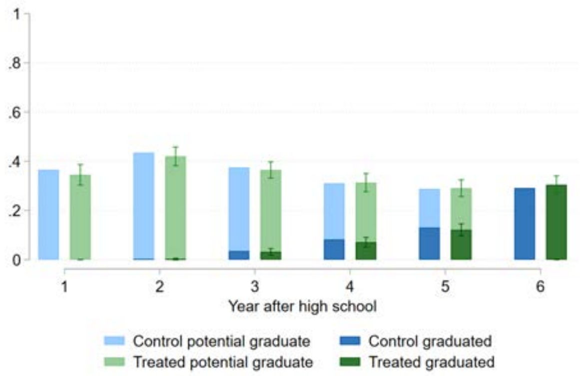
(a) Top 15% RCT sample, *girls*



(b) Top 15% RCT sample, *boys*



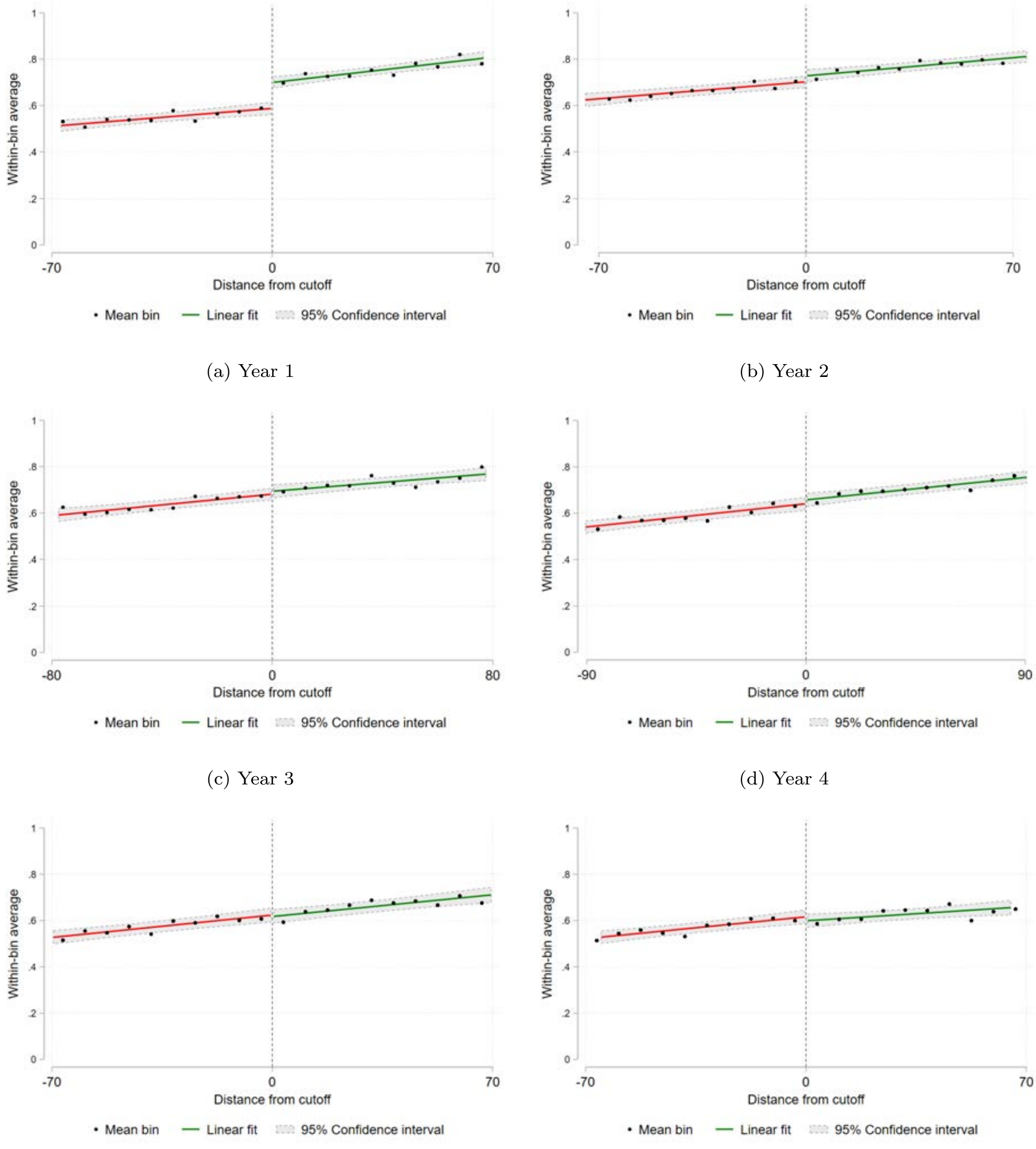
(c) Bottom 85% RCT sample, *girls*



(d) Bottom 85% RCT sample, *boys*

Notes: Define a potential graduate as a student who is enrolled in or who has graduated from higher education in the current year or in a previous year. This figure plots, for every year after high school, the fraction of top 15% and bottom 85% students at baseline in control and treated schools that are potential graduates of higher education (light shade) or that have graduated from higher education (dark shade), by gender. The error bars represent 95% confidence intervals for the difference in proportions across treatment groups (i.e., the treatment effect). Graduation data in the sixth year is not yet available for the RCT sample; the sixth-year bars represent the fraction of students who are either still enrolled in the sixth year or who have graduated during the prior five years.

Figure A7: Intent to Treat Effects on potential graduates by year post high school.



(a) Year 1

(b) Year 2

(c) Year 3

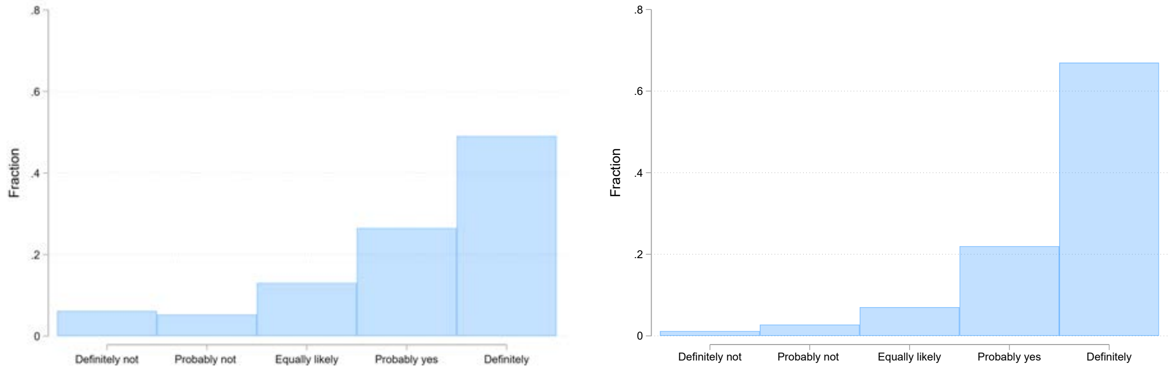
(d) Year 4

(e) Year 5

(f) Year 6

Notes: Each panel shows the average value of the corresponding outcome variable conditional on the distance of the GPA score (PRN) from the cutoff. The circles are averages across 10 equally-sized bins on either side of the cutoff, while the solid and dashed lines represent the predicted values and 95% confidence intervals using standard errors clustered at high school level. The range of the score deviation corresponds to the optimal bandwidth computed following Calonico, Cattaneo, and Titiunik (2014).

Figure A8: Beliefs about graduation chances.

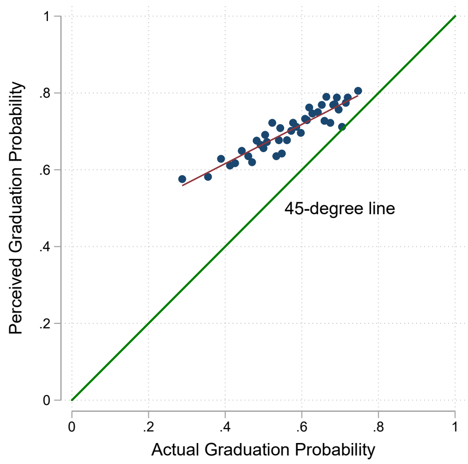


(a) All Students

(b) Enrolled in Selective College

Notes: This figure shows the histogram of high school seniors' perceived probability of graduating from a selective college. Students were asked to assess their likelihood of graduating from a selective college if they were to enroll in one. We administered the survey question to 5,809 12^{th}-grade students in 2017. The left panel shows the distribution among all survey respondents, the right panel among respondents who eventually enrolled in selective college.

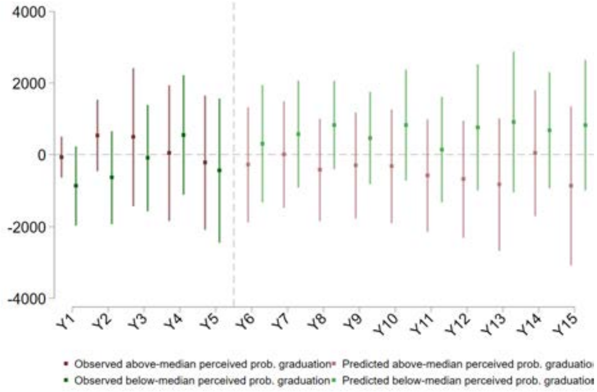
Figure A9: Perceived and actual graduation chances.



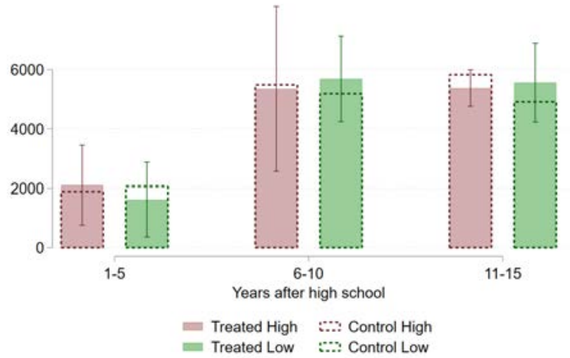
Notes: This figure is based on students who were surveyed on their beliefs regarding selective college graduation. It presents a binscatter plot comparing students' perceived graduation chances (from a survey in their final high school year) with their actual chances of having graduated or being on track to graduate six years later (from linked administrative data). Students' perceived chances were assigned values of 0, 0.25, 0.50 (for equally likely and, conservatively, for probably yes), or 1 based on their survey responses. The actual probabilities were predicted using a LASSO Probit model estimated on the sample of students who enrolled in selective colleges during the six years post high school. The data is grouped into 40 equal bins, each representing 2.5% of the total sample of 5,770 students with non-missing survey answers and LASSO-based predictions, and containing students with similar actual graduation probabilities. Each dot on the plot shows the average perceived graduation probability for the students in that bin. A 45-degree reference line is included; deviations from this line indicate discrepancies between students' beliefs and actual outcomes.

Figure A10: Additional heterogeneity analysis of labor market effects – RDD sample

Panel A: By perceived graduation probability



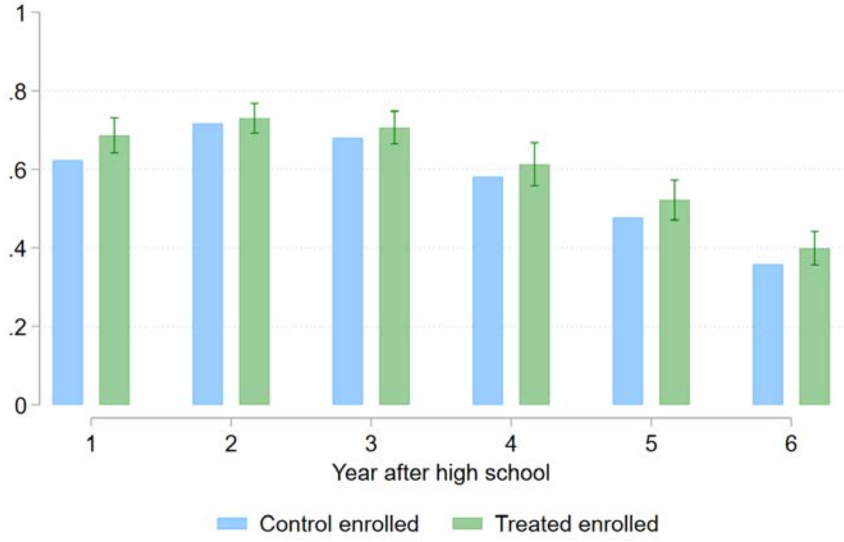
(a) Earnings, in a given year



(b) Earnings, 5-year average

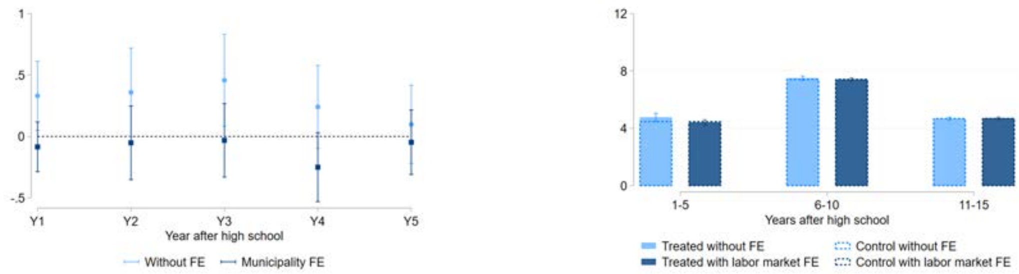
Notes: This figure is based on data from the RDD sample of students around the preferential-admission cutoff in terms of the GPA score (PRN, defined in section 2.2). The left-hand-side graphs plot estimates of parameter δ in equation (2) over time, while the right-hand-side graphs represent δ as the difference between the outcome mean just below ("control") and just above ("treated") the cutoff for a preferential admission. The x-axis indicates the year since high school graduation. Panel A shows estimates for the sub-samples of students with perceived probability of graduation above and below the median in the estimation sample. The outcome variables are observed in years 1-5, and based on the surrogate index in years 6-15 (Appendix C.2). The bars represent 95% confidence intervals for δ calculated from standard errors clustered at school level. For the outcomes based on the surrogate index, the standard errors are bootstrapped using 100 replications and resampling at school level. In the left-hand-side graphs, the dependent variable is earnings in a given year. In the right-hand-side graphs, the dependent variable is the average yearly earnings over the periods indicated on the x-axis. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). In Table A48, Panel A shows the estimated coefficients for Graph (a), Panel B for Graph (b).

Figure A11: Effects of PACE on higher education enrollment over time, RCT analysis.



Notes: This figure plots, for every year after high school, the fraction of top 15% students at baseline in control and treated schools that are enrolled in higher education. The error bars represent 95% confidence intervals for the difference in proportions across treatment groups (i.e., the treatment effect).

Figure A12: Employment effects for bottom 85% with local labor market fixed effects — experimental sample



(a) Months employed, in a given year

(b) Months employed, 5-year average

Notes: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the bottom 85% of their baseline GPA in their school. The standard errors are clustered at school level. In Graph (a) the months of employment in the year shown on the x-axis. In Graph (b) the dependent variable is the average yearly months of employment between year 1 and the year shown on the x-axis. The controls include: gender, age, socioeconomic status, mother's and father's education, family income, baseline standardized test scores, whether they ever repeated a grade, residence in Santiago, school rurality, and type of high school track (academic or vocational). The regressions include fixed effects for the school's municipality.

B Additional Tables

Table A1: Summary statistics of main outcomes, experimental and RD samples

	Top 15%			RD			Bottom 85%		
	Obs.	Mean	S.D.	Obs.	Mean	S.D.	Obs.	Mean	S.D.
A. Labor market									
Avg earnings years 1-5	2377	1713.32	2218.63	14681	1739.56	2318.08	11324	2355.01	2397.73
Avg earnings years 6-10	2437	5524.47	3123.94	15103	5254.09	3143.24	11916	5273.08	3043.01
Avg earnings years 11-15	2437	5757.63	2799.03	15103	5016.53	2842.79	11916	4174.05	2715.06
Avg months employed years 1-5	2377	3.32	3.52	14681	3.36	3.48	11324	4.56	3.50
Avg months employed years 6-10	2437	6.94	2.71	15103	6.83	2.62	11916	7.43	2.31
Avg months employed years 11-15	2437	5.06	1.40	15103	4.64	1.43	11916	4.69	1.42
B. Higher education									
In any institution									
Ever enrolled	2437	0.87	0.34	15103	0.86	0.35	11916	0.67	0.47
Graduation	2437	0.61	0.49	15103	0.59	0.49	11916	0.39	0.49
Dropout	2437	0.26	0.44	15103	0.27	0.44	11916	0.28	0.45
In selective colleges									
Ever enrolled	2437	0.47	0.50	15103	0.43	0.50	11916	0.15	0.36
Graduation	2437	0.30	0.46	15103	0.26	0.44	11916	0.09	0.28
Dropout	2437	0.12	0.32	15103	0.12	0.33	11916	0.04	0.20

Notes: This table reports the number of observations, mean and standard deviations for the labor-market and higher-education outcomes for the experimental sample, split in top 15% and bottom 85% according to the baseline high school GPA ranking, and for the RD sample within the optimal bandwidth for the first-stage regression estimated using the method by Calonico, Cattaneo, and Titiunik (2014). Earnings are in 1000 Chilean pesos (December 2022 value). Earnings and months employed in years 6-10 and 11-15 are predicted using the surrogate index (Athey et al., 2019). *Avg months* and *Avg earnings* in years s-t are the average of yearly months employed and yearly earnings from year s to year t.

Table A2: Summary statistics of baseline characteristics and main outcomes, predictions sample

	(1) Obs.	(2) Mean	(3) S.D.
A. Baseline characteristics			
Female	308899	0.52	0.50
Age	308899	16.44	1.27
Mother's education	48310	9.85	3.08
Father's education	47640	9.62	3.66
Family income	47840	187.42	169.45
SIMCE	64932	-0.35	0.84
Never failed	308899	0.87	0.34
Santiago	308899	0.25	0.43
Rural	308899	0.05	0.22
Academic track	308899	0.32	0.47
B. Labor market outcomes			
Avg earnings years 1-5	290901	2246.12	2114.27
Avg earnings years 6-10	269950	5190.42	4054.81
Avg earnings years 11-15	226972	6239.64	5119.61
Avg months employed years 1-5	290901	5.29	3.76
Avg months employed years 6-10	269950	7.77	3.80
Avg months employed years 11-15	226972	7.00	3.42
C. Higher education outcomes			
In any institution			
Ever enrolled	308899	0.57	0.49
Graduation	308899	0.36	0.48
Dropout	308899	0.21	0.41
In selective colleges			
Ever enrolled	308899	0.14	0.34
Graduation	308899	0.09	0.28
Dropout	308899	0.03	0.18

Notes: This table reports the number of observations, mean and standard deviations for pre-determined variables and outcomes of the prediction sample. The prediction sample contains students that enter higher education in 2007-2011 and that attended the same high schools of the students in the RCT and RD samples. SIMCE is a standardized achievement test taken in 10^{th} grade. Age and education are in years. Family income is the monthly family income in 1000 Chilean pesos. The variable *very low SES* (alumno prioritario) is excluded from the pre-determined variable because not available in the data for the years 2007-2011. SIMCE, Mother's and Father's Education, Family income are available only for the cohort that entered higher education in 2009 because only that cohort was surveyed. Earnings are in 1000 Chilean pesos (December 2022 value). Labor market outcomes are observed in year 15 for the coohorts entering higher education in 2007-2008; in year 14 for the coohorts entering higher education in 2007-2009; in year 13 for the coohorts entering higher education in 2007-2010; in year 1-12 for all the cohorts. *Avg months* and *Avg earnings* in years s-t are the average of yearly months employed and yearly earnings from year s to year t.

Table A3: Sample Balance Across Treatment and Control Groups

	Female	Age	Very Low SES	Mother education	Father education	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track
						Top 15%					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment	0.001	0.049	-0.020	0.129	-0.019	5.756	0.084	-0.014	0.041	-0.013	0.075
	(0.055)	(0.051)	(0.025)	(0.177)	(0.229)	(12.545)	(0.121)	(0.018)	(0.066)	(0.019)	(0.072)
p-value	0.979	0.340	0.418	0.468	0.935	0.647	0.487	0.444	0.533	0.497	0.297
Mean	0.561	16.303	0.596	9.642	9.508	282.134	-0.041	0.941	0.155	0.043	0.281
S.d.	0.496	0.587	0.491	3.132	3.103	198.181	0.805	0.237	0.362	0.203	0.450
N	2437	2437	2437	1914	1795	1919	2432	2437	2437	2437	2437
						Bottom 85%					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Treatment	-0.020	0.015	0.017	0.106	0.148	17.469	0.075	-0.013	0.040	-0.011	0.021
	(0.048)	(0.039)	(0.017)	(0.161)	(0.171)	(11.170)	(0.088)	(0.018)	(0.065)	(0.017)	(0.061)
p-value	0.672	0.699	0.322	0.512	0.385	0.119	0.394	0.478	0.533	0.534	0.737
Mean	0.479	16.598	0.609	9.495	9.273	283.378	-0.710	0.817	0.149	0.037	0.249
S.d.	0.500	0.792	0.488	3.118	3.227	205.326	0.680	0.387	0.356	0.189	0.432
N	11916	11916	11916	7754	7362	7782	11875	11916	11916	11916	11916

Notes: In this table we regress pre-determined variables on the treatment status of the top 15% and bottom 85% students according to the baseline high-school GPA ranking. Treatment is the coefficient of each regression. Standard errors clustered at the school level are shown in parentheses. The p-value is the p-value of the test of significance of the treatment coefficient. Mean and S.d. are the average and standard deviation of the pre-determined variable in the control group. Low-SES student is a student that the Government classified as very socioeconomically vulnerable (*Prioritario*). SIMCE is a standardized achievement test taken in 10^{th} grade. Age and education are in years. Family income is the monthly family income in 1000 Chilean pesos. *** p<0.01, ** p<0.05, * p<0.10.

Table A4: Tests for discontinuities in pre-determined variables

	Female	Age	Very Low SES	Mother educ.	Father educ.	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Over-conf.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Conventional	0.005	0.069***	0.007	0.144	0.064	12.000	-0.017	-0.009	0.020	-0.004	-0.025	-0.044
	(0.026)	(0.026)	(0.022)	(0.158)	(0.169)	(14.542)	(0.056)	(0.010)	(0.037)	(0.016)	(0.042)	(0.118)
Robust	0.009	0.078***	0.005	0.191	0.021	14.897	-0.028	-0.010	0.028	-0.002	-0.042	-0.084
	(0.029)	(0.028)	(0.025)	(0.173)	(0.187)	(16.514)	(0.059)	(0.011)	(0.038)	(0.017)	(0.043)	(0.133)
Bandwidth	66.868	63.101	72.524	70.534	72.382	81.185	71.260	91.547	72.184	85.097	66.070	60.273
Bandwidth N	13264	12597	14430	9182	8853	10500	14015	18101	14328	16867	13173	578
R-squared	0.000	0.006	0.000	0.001	0.001	0.000	0.038	0.008	0.001	0.000	0.005	0.004
Mean	0.595	16.261	0.610	9.993	9.732	305.467	-0.168	0.930	0.214	0.051	0.390	0.164

Notes: In this table we report the estimate for coefficient ϕ in regression equation (3), using pre-determing variables as the dependent variable. Standard errors clustered at the school level are shown in parentheses. *Mean* is the mean of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Low-SES student is a student that the Government classified as very socioeconomically vulnerable (*Prioritario*). SIMCE is a standardized achievement test taken in 10^{th} grade. Age and education are in years. Family income is the monthly family income in 1000 Chilean pesos. *** p<0.01, ** p<0.05, * p<0.10.

Table A5: ATE on earnings (RCT analysis, top 15% sample) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	-7.641	-65.108	-128.541	122.957	-53.855	220.640	-358.601
	(87.151)	(113.596)	(124.774)	(186.301)	(218.159)	(149.799)	(674.241)
Mean	797.598	1,221.364	1,514.981	2,309.703	3,330.317	8,845.511	13445.327
Total obs.	2,003	2,169	2,135	2,092	2,037	2,432	2,432
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	-7.641	-34.772	-75.516	-41.261	-27.305	11.282	222.515*
	(87.151)	(96.528)	(100.153)	(112.154)	(125.904)	(146.102)	(133.870)
Mean	797.598	999.845	1,164.598	1,430.669	1,759.669	5,571.390	5,689.375
Total obs.	2,003	2,272	2,327	2,353	2,372	2,432	2,432

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A6: ATE on months of employment (RCT analysis, top 15% sample) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	-0.156	-0.071	-0.168	0.155	-0.338	-0.032	0.048
	(0.182)	(0.222)	(0.201)	(0.245)	(0.293)	(0.050)	(0.063)
Mean	2.045	2.711	2.999	4.228	5.623	9.556	10.417
Total obs.	2,003	2,169	2,135	2,092	2,037	2,432	2,432
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	-0.156	-0.106	-0.147	-0.089	-0.101	-0.170	0.010
	(0.182)	(0.191)	(0.181)	(0.181)	(0.191)	(0.133)	(0.053)
Mean	2.045	2.349	2.563	2.954	3.410	7.054	5.061
Total obs.	2,003	2,272	2,327	2,353	2,372	2,432	2,432

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A7: ATE on being in data (RCT analysis, top 15% sample) *without controls.*

	(1)	(2)	(3)	(4)	(5)
	Panel A. In a given year				
	Year 1	Year 2	Year 3	Year 4	Year 5
Treatment	0.035*	0.003	0.019	0.018	0.035*
	(0.020)	(0.016)	(0.015)	(0.020)	(0.019)
Mean	0.812	0.890	0.872	0.854	0.826
Total obs.	2,437	2,437	2,437	2,437	2,437
	Panel A. In a given year				
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Treatment	0.035*	0.004	-0.004	-0.004	-0.002
	(0.020)	(0.012)	(0.010)	(0.009)	(0.007)
Mean	0.812	0.933	0.958	0.969	0.976
Total obs.	2,437	2,437	2,437	2,437	2,437

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is a dummy equal to 1 if the student is observed in the education or labor market dataset in the year shown in the column heading. In Panel B the dependent variable is a dummy equal to 1 if the student is observed in the education or labor market dataset at least once in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A8: ATE on earnings (RCT analysis, top 15% sample) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment – Girls	-21.505	-28.489	-61.291	-5.309	-340.801	396.717**	515.980*
	(89.158)	(118.795)	(132.934)	(179.729)	(235.613)	(192.214)	(265.040)
Mean Girls	602.476	891.624	1,031.074	1,767.288	2,701.715	8,031.507	12672.454
Total obs.	1,118	1,222	1,204	1,166	1,124	1,368	1,368
Treatment – Boys	-8.916	-143.189	-250.591	252.123	251.737	-37.489	-1497.121
	(146.291)	(182.512)	(201.194)	(283.500)	(311.190)	(225.167)	(1547.145)
Mean Boys	1,039.063	1,647.903	2,119.127	2,971.396	4,071.827	9,879.254	14424.754
Total obs.	885	947	931	926	913	1,064	1,064
p-value Girls=Boys	0.938	0.561	0.390	0.373	0.092	0.102	0.170
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment – Girls	-21.505	-33.642	-37.086	-27.130	-67.816	17.917	362.121**
	(89.158)	(92.399)	(96.665)	(107.404)	(125.585)	(187.240)	(157.367)
Mean Girls	602.476	743.085	833.085	1,054.130	1,342.039	4,732.527	4,214.020
Total obs.	1,118	1,280	1,306	1,321	1,331	1,368	1,368
Treatment – Boys	-8.916	-60.308	-146.958	-83.714	-5.724	-37.503	17.596
	(146.291)	(156.665)	(161.104)	(174.763)	(187.915)	(249.923)	(197.244)
Mean Boys	1,039.063	1,331.890	1,583.723	1,905.961	2,287.256	6,639.985	7,572.689
Total obs.	885	992	1,021	1,032	1,041	1,064	1,064
p-value Girls=Boys	0.938	0.869	0.504	0.746	0.747	0.844	0.106

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A9: ATE on education outcomes by higher-education institution *without controls* (RCT, baseline top 15%)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Ever enrolled	Potential Graduation	Dropout	Ever enrolled	Potential Graduation	Dropout
	A. Main results					
Treatment	0.140***	0.081**	0.050***	0.018	0.014	0.004
	(0.045)	(0.040)	(0.015)	(0.021)	(0.029)	(0.019)
Total obs.	2437	2437	2437	2437	2437	2437
Mean	0.423	0.278	0.103	0.860	0.604	0.257

Notes: In this table, we report the estimate for coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. *Mean* is the average of the outcome variable in the control group.

Table A10: ATE on months of employment (RCT analysis, top 15% sample) *without controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A. In a given year				
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	-0.246	-0.221	-0.375	-0.142	-0.672	-0.079	0.027
	(0.238)	(0.301)	(0.290)	(0.373)	(0.441)	(0.074)	(0.072)
Mean	2.052	2.716	2.999	4.232	5.622	9.556	10.417
Total obs.	2,007	2,172	2,140	2,095	2,041	2,437	2,437
			Panel A. In a given year				
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	-0.246	-0.227	-0.291	-0.252	-0.277	-0.336	-0.016
	(0.238)	(0.257)	(0.253)	(0.264)	(0.288)	(0.216)	(0.141)
Mean	2.052	2.355	2.567	2.958	3.412	7.054	5.061
Total obs.	2,007	2,276	2,332	2,358	2,377	2,437	2,437

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A11: ATE on earnings (RCT analysis, top 15% sample) *without controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A. In a given year				
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	-56.849	-146.528	-252.525	-87.610	-303.381	187.013	-339.576
	(114.700)	(155.037)	(173.762)	(275.979)	(340.311)	(213.826)	(730.069)
Mean	798.858	1,223.700	1,513.549	2,310.131	3,327.820	8,842.160	13441.233
Total obs.	2,007	2,172	2,140	2,095	2,041	2,437	2,437
			Panel A. In a given year				
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	-56.849	-99.206	-156.171	-141.361	-140.806	-135.955	212.132
	(114.700)	(130.707)	(137.734)	(160.857)	(188.537)	(229.309)	(261.930)
Mean	798.858	1,001.509	1,165.209	1,431.285	1,759.762	5,569.378	5,687.553
Total obs.	2,007	2,276	2,332	2,358	2,377	2,437	2,437

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A12: ATE on months of employment (RCT analysis, top 15% sample) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment – Girls	-0.193	-0.122	-0.128	-0.001	-0.678*	-0.016	0.023
	(0.202)	(0.264)	(0.241)	(0.274)	(0.345)	(0.065)	(0.066)
Mean Girls	1.723	2.272	2.330	3.571	5.044	9.312	10.294
Total obs.	1,118	1,222	1,204	1,166	1,124	1,368	1,368
Treatment – Boys	-0.128	-0.046	-0.257	0.330	0.067	-0.065	0.067
	(0.302)	(0.328)	(0.308)	(0.356)	(0.368)	(0.075)	(0.102)
Mean Boys	2.455	3.284	3.838	5.037	6.308	9.867	10.573
Total obs.	885	947	931	926	913	1,064	1,064
p-value Girls=Boys	0.853	0.847	0.728	0.408	0.080	0.630	0.696
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment – Girls	-0.193	-0.179	-0.153	-0.101	-0.174	-0.256	0.031
	(0.202)	(0.205)	(0.198)	(0.197)	(0.207)	(0.175)	(0.056)
Mean Girls	1.723	1.981	2.092	2.446	2.897	6.558	4.079
Total obs.	1,118	1,280	1,306	1,321	1,331	1,368	1,368
Treatment – Boys	-0.128	-0.042	-0.163	-0.097	-0.030	-0.080	-0.028
	(0.302)	(0.297)	(0.283)	(0.280)	(0.278)	(0.187)	(0.087)
Mean Boys	2.455	2.835	3.165	3.601	4.064	7.688	6.318
Total obs.	885	992	1,021	1,032	1,041	1,064	1,064
p-value Girls=Boys	0.853	0.677	0.975	0.989	0.634	0.413	0.538

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A13: ATE on being in data (RCT analysis, top 15% sample) *without controls*, by gender.

	(1)	(2)	(3)	(4)	(5)
Panel A. Heterogeneity by gender, in a given year					
	Year 1	Year 2	Year 3	Year 4	Year 5
Treatment – Girls	0.065***	0.009	0.047**	0.052**	0.068***
	(0.024)	(0.020)	(0.019)	(0.025)	(0.026)
Mean Girls	0.796	0.890	0.865	0.835	0.799
Total obs.	1,369	1,369	1,369	1,369	1,369
Treatment – Boys	-0.003	-0.005	-0.018	-0.025	-0.008
	(0.030)	(0.022)	(0.022)	(0.028)	(0.025)
Mean Boys	0.832	0.891	0.881	0.877	0.860
Total obs.	1,068	1,068	1,068	1,068	1,068
p-value Girls=Boys	0.060	0.634	0.023	0.026	0.029
Panel B. Heterogeneity by gender, average across years					
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Treatment – Girls	0.065***	0.010	0.008	0.012	0.007
	(0.024)	(0.016)	(0.012)	(0.011)	(0.009)
Mean Girls	0.796	0.932	0.952	0.962	0.971
Total obs.	1,369	1,369	1,369	1,369	1,369
Treatment – Boys	-0.003	-0.004	-0.020	-0.023*	-0.014
	(0.030)	(0.017)	(0.014)	(0.013)	(0.011)
Mean Boys	0.832	0.933	0.966	0.978	0.983
Total obs.	1,068	1,068	1,068	1,068	1,068
p-value Girls=Boys	0.060	0.519	0.101	0.035	0.122

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the top 15% of their school based on baseline GPA. The regressions are without controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is being in data in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is being in data in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A14: Lee bounds on selectivity and rank effects in higher-education course (RCT, baseline top 15%)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Treatment	0.182***	0.235***	0.080*	0.151***	-0.052***	-0.034**
	(0.036)	(0.034)	(0.043)	(0.044)	(0.017)	(0.016)
Total obs.	2437	2437	2437	2437	2437	2437
Selected obs.	2111	2111	2108	2108	2108	2108
Mean	0.033	0.033	0.012	0.012	0.458	0.458

Notes: In this table we report the Lee bounds for the estimate of the coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. *Total obs.* are the number of observations before the trimming procedure. *Selected obs.* are the number of observations after the trimming procedure. *Selectivity* represents the average baseline ability of peers in the first degree program a student enrolls in. Ability distance is the difference between selectivity and own baseline ability (in these regressions we do not control for own baseline ability). Rank denotes a student's relative ability among these peers: 0 if the student is the lowest-ability one and 1 if the student is the highest-ability one. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample. *** p<0.01, ** p<0.05, * p<0.10.

Table A15: Lee bounds on selectivity and rank effects in higher-education course (RCT, Top 15 % female students)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Treatment	0.213***	0.319***	0.089*	0.236***	-0.085***	-0.041**
	(0.044)	(0.043)	(0.050)	(0.050)	(0.021)	(0.020)
Total obs.	1369	1369	1369	1369	1369	1369
Selected obs.	1221	1221	1220	1220	1220	1220
Mean	0.024	0.024	-0.001	-0.001	0.468	0.468

Notes: In this table we report the Lee bounds for the estimate of the coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. *Total obs.* are the number of observations before the trimming procedure. *Selected obs.* are the number of observations after the trimming procedure. *Selectivity* represents the average baseline ability of peers in the first degree program a student enrolls in. Ability distance is the difference between selectivity and own baseline ability (in these regressions we do not control for own baseline ability). Rank denotes a student's relative ability among these peers: 0 if the student is the lowest-ability one and 1 if the student is the highest-ability one. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample. *** p<0.01, ** p<0.05, * p<0.10.

Table A16: Lee bounds on selectivity and rank effects in higher-education course (RCT, Top 15 % male students)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Treatment	0.115**	0.149**	0.011	0.083	-0.026	-0.007
	(0.054)	(0.062)	(0.074)	(0.071)	(0.025)	(0.027)
Total obs.	1068	1068	1068	1068	1068	1068
Selected obs.	890	890	888	888	888	888
Mean	0.045	0.045	0.031	0.031	0.445	0.445

Notes: In this table we report the Lee bounds for the estimate of the coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. *Total obs.* are the number of observations before the trimming procedure. *Selected obs.* are the number of observations after the trimming procedure. *Selectivity* represents the average baseline ability of peers in the first degree program a student enrolls in. Ability distance is the difference between selectivity and own baseline ability (in these regressions we do not control for own baseline ability). Rank denotes a student's relative ability among these peers: 0 if the student is the lowest-ability one and 1 if the student is the highest-ability one. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample. *** p<0.01, ** p<0.05, * p<0.10.

Table A17: LATE on being in data (RD) *without controls*.

	(1)	(2)	(3)	(4)	(5)
	Panel A. In a given year				
	Year 1	Year 2	Year 3	Year 4	Year 5
Conventional	0.064***	0.025**	0.007	0.011	0.008
	(0.014)	(0.012)	(0.015)	(0.015)	(0.016)
Robust	0.061***	0.025*	0.005	0.013	0.005
	(0.016)	(0.014)	(0.017)	(0.018)	(0.018)
Bandwidth	83.122	82.972	63.295	85.384	79.905
Bandwidth obs.	16,443	16,371	12,597	16,867	15,761
Mean	0.805	0.889	0.881	0.846	0.828
	Panel B. average across years				
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Conventional	0.064***	0.025**	0.011	0.006	0.008
	(0.014)	(0.010)	(0.008)	(0.007)	(0.006)
Robust	0.061***	0.029***	0.011	0.007	0.010
	(0.016)	(0.011)	(0.010)	(0.008)	(0.007)
Bandwidth	83.122	65.204	67.435	69.836	65.480
Bandwidth obs.	16,443	12,969	13,349	13,838	12,969
Mean	0.805	0.933	0.958	0.969	0.974

Notes: This table shows the estimate of parameter δ in equation (2). In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the being in data in the year shown in the column heading. In Panel the dependent variable is the average yearly being in data in the interval of years shown in the column heading. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A18: LATE on months of employment (RD) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A. In a given year			
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Conventional	-0.517***	-0.303*	-0.249	-0.409**	-0.333	130.549	130.549
	(0.147)	(0.175)	(0.179)	(0.207)	(0.209)	(103.314)	(103.314)
Robust	-0.554***	-0.268	-0.223	-0.442*	-0.328	116.988	116.988
	(0.169)	(0.199)	(0.204)	(0.239)	(0.244)	(122.810)	(122.810)
Bandwidth	71.211	84.087	83.134	65.485	80.943	76.199	76.199
Bandwidth obs.	11,600	14,680	14,167	10,899	13,097	14,972	14,972
Mean	2.231	2.894	3.206	4.135	5.091	5,044.060	5,044.060
				Panel B. average across years			
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Conventional	-0.517***	-0.366***	-0.308**	-0.345**	-0.346**	130.549	130.549
	(0.147)	(0.137)	(0.139)	(0.142)	(0.140)	(103.314)	(103.314)
Robust	-0.554***	-0.354**	-0.285*	-0.336**	-0.334**	116.988	116.988
	(0.169)	(0.153)	(0.155)	(0.161)	(0.159)	(122.810)	(122.810)
Bandwidth	71.211	98.442	93.755	86.754	84.451	76.199	76.199
Bandwidth obs.	11,600	18,024	17,521	16,450	16,039	14,972	14,972
Mean	2.231	2.528	2.729	3.007	3.362	5,044.060	5,044.060

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of employment in the interval of years shown in the column heading. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A19: LATE on earnings (RD) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Conventional	-212.231***	-163.741	-132.770	-302.966**	-344.806**	130.549	130.549
	(67.306)	(99.816)	(112.227)	(134.613)	(168.448)	(103.314)	(103.314)
Robust	-199.849***	-140.950	-104.560	-281.547*	-305.978	116.988	116.988
	(75.786)	(111.047)	(125.384)	(153.290)	(188.935)	(122.810)	(122.810)
Bandwidth	83.125	86.937	86.607	80.900	84.778	76.199	76.199
Bandwidth obs.	13,463	15,162	14,800	13,383	13,740	14,972	14,972
Mean	896.254	1,309.300	1,601.545	2,287.678	3,163.065	5,044.060	5,044.060
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Conventional	-212.231***	-145.444*	-143.636*	-164.798*	-173.021*	130.549	130.549
	(67.306)	(75.057)	(82.633)	(85.810)	(93.628)	(103.314)	(103.314)
Robust	-199.849***	-125.157	-118.357	-144.288	-155.032	116.988	116.988
	(75.786)	(82.803)	(90.224)	(95.251)	(103.860)	(122.810)	(122.810)
Bandwidth	83.125	89.676	84.409	91.052	90.127	76.199	76.199
Bandwidth obs.	13,463	16,434	15,771	17,262	17,181	14,972	14,972
Mean	896.254	1,066.667	1,227.445	1,471.844	1,731.744	5,044.060	5,044.060

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of employment in the interval of years shown in the column heading. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A20: LATE on earnings (RD) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
				Girls			
Conventional	-180.463**	70.076	234.041	-81.672	-80.211	298.011**	388.497**
	(71.129)	(124.190)	(160.035)	(165.513)	(226.729)	(134.819)	(193.004)
Robust	-160.348**	107.439	303.071*	-41.522	8.150	310.692*	340.624
	(79.794)	(141.825)	(177.107)	(191.039)	(253.085)	(159.772)	(224.453)
Bandwidth	83.327	58.298	43.076	62.087	51.606	62.719	78.409
Bandwidth obs.	7,711	6,057	4,525	6,083	4,942	7,353	9,097
Mean	755.460	1,026.909	1,235.885	1,808.708	2,642.963	7,551.148	12071.631
				Boys			
Conventional	-247.101**	-266.573	-375.623*	-584.100**	-375.043	-113.106	-74.124
	(116.041)	(169.072)	(225.561)	(257.544)	(318.481)	(203.858)	(345.809)
Robust	-260.287*	-245.786	-370.727	-653.105**	-433.916	-165.286	-160.061
	(133.643)	(195.032)	(263.856)	(294.883)	(368.872)	(221.674)	(387.107)
Bandwidth	82.470	80.595	63.125	59.876	58.331	69.254	64.176
Bandwidth obs.	5,692	5,843	4,477	4,187	4,042	5,578	5,149
Mean	1,063.321	1,611.702	2,096.992	2,943.149	3,853.829	9,724.447	14561.635
p-value Girls=Boys	0.592	0.080	0.030	0.085	0.424	0.119	0.141
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
				Girls			
Conventional	-180.463**	-72.383	107.974	33.242	-31.240	122.831	250.451**
	(71.129)	(88.428)	(106.217)	(110.269)	(116.005)	(165.893)	(109.759)
Robust	-160.348**	-68.539	146.480	70.402	7.309	162.841	224.685*
	(79.794)	(104.352)	(118.939)	(124.904)	(130.810)	(192.276)	(131.849)
Bandwidth	83.327	72.025	51.410	52.533	53.878	50.578	85.574
Bandwidth obs.	7,711	7,819	5,796	6,021	6,166	5,993	9,921
Mean	755.460	892.817	959.122	1,151.593	1,420.898	4,544.234	3,548.290
				Boys			
Conventional	-247.101**	-212.601*	-232.891	-240.319	-266.456	-266.277	-20.749
	(116.041)	(128.157)	(146.226)	(160.775)	(175.819)	(259.734)	(174.206)
Robust	-260.287*	-208.708	-249.378	-236.811	-259.056	-294.689	-45.614
	(133.643)	(148.141)	(167.571)	(186.632)	(204.144)	(299.666)	(192.192)
Bandwidth	82.470	81.452	75.375	68.820	65.455	66.533	75.073
Bandwidth obs.	5,692	6,162	5,777	5,343	5,088	5,359	6,042
Mean	1,063.321	1,314.843	1,457.769	1,828.729	2,165.498	6,360.976	7,242.711
p-value Girls=Boys	0.592	0.307	0.045	0.145	0.250	0.127	0.089

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the earnings in the year shown in the column heading for the subsamples of female and male, respectively. In Panels B the dependent variable is the average yearly earnings in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A21: LATE on earnings (RD) *with controls*, by overconfidence.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by overconfidence, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
	High overconfidence						
Conventional	124.104	745.282	987.516	457.567	-289.426	-351.155	-1277.129
	(315.184)	(506.758)	(1083.536)	(1069.090)	(1145.589)	(783.818)	(1366.443)
Robust	203.900	870.796	1223.791	697.087	-113.992	-481.245	-1041.347
	(368.037)	(574.364)	(1309.498)	(1289.413)	(1326.421)	(939.491)	(1640.646)
Bandwidth	65.140	72.460	53.662	61.596	72.000	57.271	30.360
Bandwidth obs.	303	381	287	313	363	329	173
Mean	703.923	529.173	1,147.534	2,515.854	3,478.306	8,891.218	13581.463
	Low overconfidence						
Conventional	-847.831*	-842.607	-764.779	146.676	-1764.359*	180.570	265.768
	(471.441)	(675.068)	(724.201)	(794.751)	(1013.816)	(789.611)	(1051.146)
Robust	-928.336*	-847.494	-941.811	145.907	-1933.257*	273.632	162.762
	(548.920)	(798.149)	(822.402)	(928.942)	(1140.199)	(857.438)	(1182.281)
Bandwidth	48.908	49.553	67.200	92.585	77.099	68.059	65.591
Bandwidth obs.	154	162	219	298	235	261	247
Mean	1,557.660	1,775.534	1,392.086	2,785.033	4,328.509	7,733.705	12167.470
p-value High=Low	0.068	0.028	0.151	0.812	0.347	0.508	0.361
	Panel B. Heterogeneity by overconfidence, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
	High overconfidence						
Conventional	124.104	347.727	708.147	594.035	373.372	-73.441	-471.715
	(315.184)	(390.167)	(622.212)	(675.170)	(693.449)	(774.839)	(503.496)
Robust	203.900	425.530	810.877	740.681	499.524	-84.633	-619.377
	(368.037)	(438.480)	(747.271)	(804.555)	(836.427)	(902.453)	(580.263)
Bandwidth	65.140	82.583	59.704	59.866	62.497	62.786	44.854
Bandwidth obs.	303	445	336	338	353	356	261
Mean	703.923	842.163	971.406	1,382.023	1,723.021	5,367.345	5,820.774
	Low overconfidence						
Conventional	-847.831*	-909.619*	-774.717	-791.771	-872.469	24.667	415.200
	(471.441)	(516.737)	(520.042)	(515.230)	(579.528)	(853.035)	(647.221)
Robust	-928.336*	-934.946	-908.977	-919.971	-1103.688*	155.220	489.408
	(548.920)	(617.028)	(610.245)	(601.495)	(640.059)	(926.556)	(717.656)
Bandwidth	48.908	47.400	52.116	64.557	59.179	59.769	66.542
Bandwidth obs.	154	166	189	233	223	232	252
Mean	1,557.660	1,600.552	1,520.919	1,673.323	2,219.927	5,180.826	4,451.687
p-value High=Low	0.068	0.021	0.061	0.064	0.150	0.912	0.199

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the earnings in the year shown in the column heading for the subsamples of above median and below median overconfidence, respectively. In Panels B the dependent variable is the average yearly earnings in the interval of years shown in the column heading for the subsamples of above median and below median overconfidence, respectively. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A22: LATE on education outcomes by higher-education institution *without controls* (RD)

	(1)	(2)	(3)	(4)	(5)	(6)
		Selective College			Any Institution	
			Results without controls			
	Ever enrolled	Potential Graduation	Dropout	Ever enrolled	Potential Graduation	Dropout
Conventional	0.164***	0.070**	0.096***	0.041***	-0.021	0.052**
	(0.034)	(0.030)	(0.014)	(0.015)	(0.026)	(0.021)
Robust	0.150***	0.060*	0.097***	0.038**	-0.030	0.059**
	(0.035)	(0.031)	(0.016)	(0.018)	(0.029)	(0.023)
Total obs.	54248	54248	54248	54248	54248	54248
Bandwidth	57.830	82.852	76.529	81.218	65.172	66.309
Bandwidth obs.	11518	16371	15191	16077	12969	13173
R-squared	0.110	0.046	0.036	0.020	0.003	0.002
Mean	0.397	0.248	0.088	0.851	0.616	0.241
			Results with controls			
	Ever enrolled	Potential Graduation	Dropout	Ever enrolled	Potential Graduation	Dropout
Conventional	0.171***	0.055***	0.096***	0.041***	-0.023	0.065***
	(0.024)	(0.021)	(0.014)	(0.014)	(0.023)	(0.021)
Robust	0.163***	0.048**	0.096***	0.044***	-0.031	0.072***
	(0.026)	(0.023)	(0.017)	(0.017)	(0.026)	(0.023)
Total obs.	54248	54248	54248	54248	54248	54248
Bandwidth	53.116	63.887	74.663	68.923	53.142	56.033
Bandwidth obs.	10645	12682	14810	13668	10645	11233
R-squared	0.104	0.037	0.036	0.019	0.001	0.002
Mean	0.402	0.259	0.090	0.857	0.618	0.234

Notes: In this table we report the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Mean* is the mean of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** p<0.01, ** p<0.05, * p<0.10.

Table A23: LATE on months of employment (RD) *without controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Conventional	-0.478***	-0.197	-0.121	-0.281	-0.162	-0.054	0.004
	(0.179)	(0.233)	(0.244)	(0.271)	(0.310)	(0.045)	(0.040)
Robust	-0.459**	-0.117	-0.057	-0.254	-0.061	-0.042	-0.008
	(0.203)	(0.251)	(0.266)	(0.301)	(0.334)	(0.054)	(0.049)
Bandwidth	68.747	77.145	72.727	67.666	66.937	92.315	81.055
Bandwidth obs.	11,314	13,591	12,548	11,383	11,012	18,217	16,077
Mean	2.231	2.894	3.206	4.135	5.091	9.379	10.320
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Conventional	-0.478***	-0.343*	-0.257	-0.195	-0.170	-0.160	-0.012
	(0.179)	(0.185)	(0.192)	(0.202)	(0.207)	(0.125)	(0.061)
Robust	-0.459**	-0.309	-0.212	-0.148	-0.109	-0.109	-0.028
	(0.203)	(0.206)	(0.209)	(0.221)	(0.223)	(0.143)	(0.073)
Bandwidth	68.747	78.807	77.427	71.443	69.970	78.029	83.724
Bandwidth obs.	11,314	14,593	14,625	13,667	13,465	15,506	16,555
Mean	2.231	2.528	2.729	3.007	3.362	6.894	4.656

Notes: This table shows the estimate of parameter δ in equation (2). In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel the dependent variable is the average yearly months of employment in the interval of years shown in the column heading. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A24: LATE on earnings (RD) *without controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A. In a given year			
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Conventional	-202.372**	-89.020	-86.572	-180.736	-211.178	119.718	47.192
	(80.900)	(128.132)	(146.938)	(183.047)	(234.839)	(127.793)	(212.766)
Robust	-202.645**	-46.222	-54.523	-136.384	-145.509	85.706	-26.849
	(92.335)	(138.656)	(161.443)	(202.135)	(254.004)	(158.405)	(248.370)
Bandwidth	81.065	76.398	75.574	70.336	71.381	92.215	55.671
Bandwidth obs.	13,283	13,411	13,014	11,788	11,735	18,217	11,099
Mean	896.254	1,309.300	1,601.545	2,287.678	3,163.065	8,421.924	13050.653
				Panel B. average across years			
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Conventional	-202.372**	-119.270	-101.331	-116.166	-92.327	-90.946	66.158
	(80.900)	(96.400)	(106.491)	(117.358)	(130.848)	(119.696)	(137.298)
Robust	-202.645**	-89.411	-75.416	-85.060	-54.051	-53.677	11.218
	(92.335)	(104.971)	(117.090)	(128.025)	(141.536)	(138.978)	(166.163)
Bandwidth	81.065	77.637	78.282	76.971	72.289	81.427	75.744
Bandwidth obs.	13,283	14,395	14,820	14,674	13,935	16,077	14,985
Mean	896.254	1,066.667	1,227.445	1,471.844	1,731.744	5,305.988	5,044.060

Notes: This table shows the estimate of parameter δ in equation (2). In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the earnings in the year shown in the column heading. In Panel the dependent variable is the average yearly earnings in the interval of years shown in the column heading. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A25: LATE on months of employment (RD) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
				Girls			
Conventional	-0.538***	-0.062	-0.004	-0.435*	-0.473*	-0.019	0.053
	(0.166)	(0.239)	(0.246)	(0.233)	(0.253)	(0.062)	(0.057)
Robust	-0.496***	-0.047	0.023	-0.462*	-0.489*	-0.015	0.049
	(0.187)	(0.278)	(0.289)	(0.271)	(0.294)	(0.076)	(0.069)
Bandwidth	80.527	67.125	64.027	86.385	92.457	68.631	72.422
Bandwidth obs.	7,500	6,894	6,496	8,307	8,556	8,010	8,398
Mean	2.101	2.692	2.850	3.791	4.806	9.145	10.162
				Boys			
Conventional	-0.533**	-0.417	-0.350	-0.594*	-0.026	-0.081	-0.067
	(0.223)	(0.259)	(0.313)	(0.317)	(0.325)	(0.077)	(0.067)
Robust	-0.545**	-0.363	-0.316	-0.618*	0.038	-0.074	-0.089
	(0.262)	(0.300)	(0.366)	(0.370)	(0.375)	(0.090)	(0.076)
Bandwidth	73.205	82.001	65.100	62.803	68.594	74.034	60.017
Bandwidth obs.	5,070	5,977	4,623	4,377	4,775	5,969	4,839
Mean	2.429	3.144	3.754	4.701	5.523	9.700	10.549
p-value Girls=Boys	0.986	0.267	0.390	0.671	0.245	0.497	0.154
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
				Girls			
Conventional	-0.538***	-0.324*	-0.145	-0.268	-0.293	-0.204	-0.017
	(0.166)	(0.180)	(0.190)	(0.186)	(0.179)	(0.139)	(0.044)
Robust	-0.496***	-0.327	-0.129	-0.293	-0.301	-0.187	-0.032
	(0.187)	(0.211)	(0.222)	(0.215)	(0.205)	(0.161)	(0.052)
Bandwidth	80.527	75.051	64.698	70.016	75.498	58.811	78.241
Bandwidth obs.	7,500	8,131	7,230	7,903	8,517	6,913	9,097
Mean	2.101	2.308	2.480	2.780	3.113	6.490	3.731
				Boys			
Conventional	-0.533**	-0.471**	-0.383*	-0.339	-0.323	-0.151	-0.006
	(0.223)	(0.211)	(0.224)	(0.233)	(0.241)	(0.173)	(0.065)
Robust	-0.545**	-0.475*	-0.362	-0.311	-0.251	-0.132	-0.006
	(0.262)	(0.247)	(0.259)	(0.270)	(0.273)	(0.198)	(0.078)
Bandwidth	73.205	79.440	75.914	66.650	63.402	64.141	99.591
Bandwidth obs.	5,070	6,004	5,812	5,174	4,927	5,149	8,062
Mean	2.429	2.716	2.954	3.415	3.828	7.393	6.035
p-value Girls=Boys	0.986	0.550	0.386	0.800	0.917	0.781	0.856

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading for the subsamples of female and male, respectively. In Panels B the dependent variable is the average yearly months of employment in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A26: LATE on being in data (RD) *without controls*, by gender.

	(1)	(2)	(3)	(4)	(5)
Panel A. Heterogeneity by gender, in a given year					
	Year 1	Year 2	Year 3	Year 4	Year 5
Girls					
Conventional	0.084***	0.017	-0.008	0.017	0.025
	(0.021)	(0.016)	(0.020)	(0.021)	(0.023)
Robust	0.079***	0.016	-0.014	0.018	0.021
	(0.024)	(0.018)	(0.023)	(0.025)	(0.026)
Bandwidth	66.844	82.396	64.398	80.442	66.823
Bandwidth obs.	7,855	9,611	7,585	9,374	7,855
Mean	0.772	0.888	0.887	0.832	0.796
Boys					
Conventional	0.052***	0.024	0.033	0.018	-0.011
	(0.019)	(0.018)	(0.022)	(0.021)	(0.022)
Robust	0.058***	0.031	0.038	0.021	-0.005
	(0.022)	(0.021)	(0.026)	(0.025)	(0.026)
Bandwidth	88.111	75.551	57.906	72.235	69.895
Bandwidth obs.	7,161	6,132	4,687	5,855	5,654
Mean	0.846	0.901	0.871	0.862	0.866
p-value Girls=Boys	0.270	0.772	0.159	0.977	0.249
Panel B. Heterogeneity by gender, average across years					
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Girls					
Conventional	0.084***	0.031**	0.016	0.017*	0.021**
	(0.021)	(0.012)	(0.010)	(0.009)	(0.008)
Robust	0.079***	0.033**	0.019	0.020*	0.024**
	(0.024)	(0.014)	(0.012)	(0.010)	(0.009)
Bandwidth	66.844	75.223	64.817	59.917	53.965
Bandwidth obs.	7,855	8,816	7,629	7,090	6,365
Mean	0.772	0.927	0.960	0.969	0.974
Boys					
Conventional	0.052***	0.020	0.002	0.001	0.001
	(0.019)	(0.015)	(0.013)	(0.011)	(0.011)
Robust	0.058***	0.021	0.003	0.004	0.003
	(0.022)	(0.018)	(0.016)	(0.013)	(0.013)
Bandwidth	88.111	70.470	70.999	78.205	72.512
Bandwidth obs.	7,161	5,690	5,735	6,332	5,883
Mean	0.846	0.938	0.960	0.966	0.974
p-value Girls=Boys	0.270	0.583	0.396	0.287	0.163

Notes: This table shows the estimate of parameter δ in equation (2). All regressions are without controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the being in data in the year shown in the column heading for the subsamples of female and male, respectively. In Panels B the dependent variable is the average yearly being in data in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

Table A27: Bounds on selectivity and rank effects in higher education course (RD, all students)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Conventional	0.153***	0.261***	0.085**	0.281***	-0.099***	-0.042**
	(0.040)	(0.042)	(0.040)	(0.037)	(0.015)	(0.017)
Robust	0.061	0.174***	0.093**	0.303***	-0.110***	-0.049**
	(0.044)	(0.044)	(0.045)	(0.045)	(0.017)	(0.019)
Bandwidth	85	85	85	85	85	85
Observations	14077	14078	14076	14077	14076	14077
Mean	-0.071	-0.071	0.069	0.069	0.437	0.437

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Selectivity is average baseline ability of student peers in the same degree program. Ability distance is the difference between selectivity and own baseline ability. Rank is the position in the baseline-ability ranking from 0 to 1 in the same degree program. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample.

Table A28: Bounds on selectivity and rank effects in higher education course (RD, Female students)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Conventional	0.115***	0.205***	0.095**	0.324***	-0.115***	-0.044**
	(0.043)	(0.042)	(0.048)	(0.047)	(0.019)	(0.021)
Robust	0.060	0.151***	0.092	0.318***	-0.118***	-0.047*
	(0.052)	(0.050)	(0.063)	(0.063)	(0.025)	(0.027)
Bandwidth	63	63	59	59	59	59
Observations	6413	6414	5989	5990	5989	5990
Mean	-0.105	-0.105	0.073	0.073	0.441	0.441

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Selectivity is average baseline ability of student peers in the same degree program. Ability distance is the difference between selectivity and own baseline ability. Rank is the position in the baseline-ability ranking from 0 to 1 in the same degree program. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample.

Table A29: Bounds on selectivity and rank effects in higher education course (RD, Male students)

	Selectivity		Ability distance		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Conventional	0.129*	0.215***	0.097	0.245***	-0.088***	-0.048*
	(0.066)	(0.072)	(0.065)	(0.059)	(0.024)	(0.026)
Robust	0.094	0.174**	0.062	0.223***	-0.082***	-0.038
	(0.072)	(0.074)	(0.072)	(0.070)	(0.029)	(0.030)
Bandwidth	70	70	69	69	69	69
Observations	4686	4687	4558	4559	4558	4559
Mean	0.007	0.007	0.034	0.034	0.444	0.444

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Selectivity is average baseline ability of student peers in the same degree program. Ability distance is the difference between selectivity and own baseline ability. Rank is the position in the baseline-ability ranking from 0 to 1 in the same degree program. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample.

Table A30: Balance Across Samples with and without Missing Overconfidence Data

	Female	Age	Very Low SES	Mother education	Father education	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track
	RD										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Missing	0.091	-0.025	0.020	-0.062	0.022	9.378	-0.002	0.008	0.057	0.020	0.020
	(0.056)	(0.037)	(0.021)	(0.173)	(0.245)	(13.275)	(0.109)	(0.013)	(0.057)	(0.019)	(0.066)
p-value	0.106	0.492	0.360	0.721	0.929	0.480	0.983	0.575	0.319	0.290	0.766
Mean	0.504	16.360	0.593	10.036	9.770	301.292	-0.210	0.914	0.163	0.032	0.350
S.d.	0.500	0.635	0.492	3.060	3.306	226.763	0.815	0.280	0.370	0.176	0.477
N	15103	15103	15103	9856	9322	9854	14972	15103	15103	15103	15103

Notes: This table shows results from regressions estimated on the RD sample of students whose PRN score is within the optimal bandwidth for the first-stage regression estimated using the method by Calonico, Cattaneo, and Titiunik (2014). We regress pre-determined variables on a dummy equal to one if data on overconfidence is missing, and to zero otherwise, and report the dummy coefficient for each regression. Standard errors clustered at the school level are shown in parentheses. The p-value is the p-value of the test of significance of the dummy coefficient. Mean and S.d. are the average and standard deviation of the pre-determined variable in the sample with non-missing overconfidence. Low-SES student is a student that the Government classified as very socioeconomically vulnerable (*Prioritario*). SIMCE is a standardized achievement test taken in 10^{th} grade. Age and education are in years. Family income is the monthly family income in 1000 Chilean pesos. *** p<0.01, ** p<0.05, * p<0.10.

Table A31: Perceived and actual chances of graduation from selective college among those who enroll.

Survey Answer	**Actual % Graduates**
Chances of Graduating <= 50%	52.26
Will Probably Graduate	59.55
Will Certainly Graduate	64.04
Any Survey Answer	61.75

Notes: The table uses the sample of students who enrolled in a selective college during the six years after high school and who were surveyed on their beliefs regarding selective college graduation conditional on enrolling. Beliefs were collected through a survey in the last high school year (2017). Information on actual college performance comes from linked administrative records for the same students six years after leaving high school (2023). Each row restricts the sample according to students' survey answers, and shows among the students who gave each answer what percentage have graduated or are on track to graduate from a selective college six years later.

Table A32: Gender gap in overconfidence

	(1) Overconfidence	(2) Overconfidence
Female	-0.283***	-0.281***
	(0.031)	(0.032)
Controls	NO	YES
Obs.	5770	5770
Mean	0.131	0.131

Notes: In this table we regress overconfidence on the gender dummy, in the sample of survey respondents. Overconfidence is the difference between the perceived and the actual likelihood of graduating from a selective college (see Appendix C.1), standardized to have mean zero and variance one. Standard errors clustered at the school level are shown in parentheses. Mean refers to average overconfidence among male students. The regression in column (2) includes the following controls: age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** p<0.01, ** p<0.05, * p<0.10.

Table A33: Bounds on selectivity and rank effects in higher education course (RD, Students above median overconfidence)

	Selectivity		Ability distance		Rank	
	(1) Lower bound	(2) Upper bound	(3) Lower bound	(4) Upper bound	(5) Lower bound	(6) Upper bound
Conventional	0.207	0.217	0.149	0.246	-0.025	-0.014
	(0.135)	(0.137)	(0.250)	(0.248)	(0.093)	(0.092)
Robust	0.026	0.080	0.709**	0.817***	-0.167	-0.160
	(0.216)	(0.209)	(0.326)	(0.317)	(0.124)	(0.124)
Bandwidth	64	64	64	64	64	64
Observations	314	313	314	313	314	313
Mean	-0.002	-0.002	0.012	0.012	0.457	0.457

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Selectivity is average baseline ability of student peers in the same degree program. Ability distance is the difference between selectivity and own baseline ability. Rank is the position in the baseline-ability ranking from 0 to 1 in the same degree program. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample.

Table A34: Bounds on selectivity and rank effects in higher education course (RD, Students below median overconfidence)

	Selectivity		Ability distance		Rank	
	(1) Lower bound	(2) Upper bound	(3) Lower bound	(4) Upper bound	(5) Lower bound	(6) Upper bound
Conventional	0.295*	0.368**	-0.021	0.258	-0.055	-0.007
	(0.162)	(0.166)	(0.193)	(0.206)	(0.085)	(0.084)
Robust	0.400**	0.476**	-0.212	0.189	0.014	0.115
	(0.197)	(0.191)	(0.257)	(0.274)	(0.110)	(0.104)
Bandwidth	102	102	102	102	102	102
Observations	304	303	304	303	304	303
Mean	-0.143	-0.143	0.058	0.058	0.447	0.447

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Standard errors clustered at the school level are shown in parentheses. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. Selectivity is average baseline ability of student peers in the same degree program. Ability distance is the difference between selectivity and own baseline ability. Rank is the position in the baseline-ability ranking from 0 to 1 in the same degree program. *Mean* is the mean of the outcome variable just below the cutoff in the untrimmed sample.

Table A35: LATE on education outcomes by higher-education institution and beliefs in graduation probability *with controls* (RD)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Ever enrolled	Graduation	Dropout	Ever enrolled	Graduation	Dropout
	Higher beliefs in graduation probability					
Conventional	0.145	-0.120	0.244***	0.034	-0.220**	0.364***
	(0.111)	(0.102)	(0.088)	(0.064)	(0.106)	(0.099)
Robust	0.121	-0.174	0.269***	0.021	-0.251**	0.395***
	(0.126)	(0.113)	(0.098)	(0.073)	(0.124)	(0.112)
Total obs.	1374	1374	1374	1374	1374	1374
Bandwidth	69.424	34.311	47.227	64.409	46.558	41.088
Bandwidth obs.	401	210	282	369	279	250
Mean	0.509	0.402	0.090	0.899	0.755	0.086
	Lower beliefs in graduation probability					
Conventional	0.184	0.143	0.100	-0.025	0.138	-0.139
	(0.151)	(0.101)	(0.118)	(0.088)	(0.134)	(0.126)
Robust	0.202	0.150	0.138	-0.038	0.147	-0.146
	(0.180)	(0.116)	(0.135)	(0.105)	(0.154)	(0.145)
Total obs.	1167	1167	1167	1167	1167	1167
Bandwidth	53.985	84.277	45.887	92.503	90.172	70.595
Bandwidth obs.	217	337	187	364	359	284
Mean	0.285	0.122	0.088	0.772	0.415	0.364
p-value High=Low	0.840	0.053	0.324	0.625	0.030	0.002

Notes: In this table we report the estimate for coefficient δ in regression equation (2) for the subsamples of students with above median and below median perceived probability of graduation in a selective college. Standard errors clustered at the school level are shown in parentheses. *Mean* is the mean of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidths, a linear polynomial of the ranking score and uniform kernels are used in all the specifications. The regressions include the following controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, and high school track (academic or vocational). *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

Table A36: ATE on months of employment (RCT analysis, bottom 85% sample) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	0.330**	0.359*	0.457**	0.240	0.098	0.010	0.005
	(0.143)	(0.184)	(0.191)	(0.172)	(0.162)	(0.034)	(0.045)
Mean	3.092	3.808	4.167	5.743	6.728	9.129	9.701
Total obs.	8,493	9,670	9,305	9,462	9,322	11,875	11,875
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	0.330**	0.344**	0.359**	0.342**	0.292**	0.084	0.002
	(0.143)	(0.158)	(0.158)	(0.153)	(0.147)	(0.078)	(0.049)
Mean	3.092	3.327	3.518	4.004	4.462	7.404	4.672
Total obs.	8,493	10,369	10,850	11,149	11,286	11,875	11,875

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A37: ATE on being in data (RCT analysis, bottom 85% sample) *without controls.*

	(1)	(2)	(3)	(4)	(5)
	Panel A. In a given year				
	Year 1	Year 2	Year 3	Year 4	Year 5
Treatment	0.033**	0.021**	0.017	0.023**	0.016
	(0.016)	(0.011)	(0.013)	(0.011)	(0.011)
Mean	0.705	0.807	0.777	0.789	0.779
Total obs.	11,916	11,916	11,916	11,916	11,916
	Panel A. In a given year				
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Treatment	0.033**	0.015	0.012	0.010	0.005
	(0.016)	(0.010)	(0.008)	(0.006)	(0.006)
Mean	0.705	0.868	0.910	0.936	0.949
Total obs.	11,916	11,916	11,916	11,916	11,916

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is a dummy equal to 1 if the student is observed in the education or labor market dataset in the year shown in the column heading. In Panel B the dependent variable is a dummy equal to 1 if the student is observed in the education or labor market dataset at least once in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A38: ATE on months of employment (RCT analysis, bottom 85% sample) *without controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	0.329	0.352	0.461	0.227	0.073	0.021	0.024
	(0.205)	(0.271)	(0.299)	(0.287)	(0.271)	(0.047)	(0.050)
Mean	3.093	3.809	4.168	5.746	6.729	9.128	9.697
Total obs.	8,524	9,702	9,332	9,492	9,349	11,916	11,916
	Panel A. In a given year						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	0.329	0.334	0.354	0.331	0.278	0.080	0.051
	(0.205)	(0.232)	(0.238)	(0.240)	(0.238)	(0.127)	(0.113)
Mean	3.093	3.327	3.519	4.005	4.463	7.403	4.670
Total obs.	8,524	10,405	10,886	11,187	11,324	11,916	11,916

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A39: ATE on earnings (RCT analysis, bottom 85% sample) *without controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	190.883*	237.893	314.371	158.882	5.940	127.635	174.887
	(107.455)	(160.893)	(201.075)	(230.914)	(260.710)	(140.117)	(194.176)
Mean	1,222.932	1,701.007	2,072.702	3,200.554	4,140.276	7,222.950	10658.389
Total obs.	8,524	9,702	9,332	9,492	9,349	11,916	11,916
	Panel A. In a given year						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	190.883*	207.107	227.501	212.789	174.097	121.514	133.617
	(107.455)	(130.165)	(144.322)	(158.416)	(171.507)	(158.949)	(189.246)
Mean	1,222.932	1,399.448	1,572.724	1,928.036	2,297.047	5,232.765	4,129.729
Total obs.	8,524	10,405	10,886	11,187	11,324	11,916	11,916

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use no controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A40: ATE on education outcomes by higher-education institution *without controls* (RCT, bottom 85%)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selective College			Any Institution		
	Ever enrolled	Potential Graduation	Dropout	Ever enrolled	Potential Graduation	Dropout
	A. Main results					
Treatment	0.030	0.012	0.012*	-0.005	-0.000	-0.005
	(0.023)	(0.016)	(0.007)	(0.024)	(0.029)	(0.014)
Total obs.	11916	11916	11916	11916	11916	11916
Mean	0.138	0.085	0.039	0.672	0.390	0.282

Notes: In this table, we report the estimate for coefficient β in regression equation (1). Standard errors clustered at the school level are shown in parentheses. *Mean* is the average of the outcome variable in the control group.

Table A41: ATE on months of employment (RCT analysis, bottom 85% sample) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment – Girls	0.171	0.145	0.411	0.205	0.239	-0.014	0.008
	(0.166)	(0.225)	(0.261)	(0.245)	(0.234)	(0.041)	(0.050)
Mean Girls	2.503	3.033	3.124	4.648	5.734	8.856	9.572
Total obs.	3,791	4,438	4,240	4,296	4,234	5,617	5,617
Treatment – Boys	0.425**	0.511**	0.471**	0.257	-0.023	0.028	0.000
	(0.179)	(0.228)	(0.204)	(0.181)	(0.169)	(0.047)	(0.058)
Mean Boys	3.584	4.491	5.073	6.692	7.584	9.379	9.813
Total obs.	4,702	5,232	5,065	5,166	5,088	6,258	6,258
p-value Girls=Boys	0.202	0.172	0.817	0.835	0.258	0.460	0.929
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment – Girls	0.171	0.147	0.221	0.215	0.215	0.067	0.000
	(0.166)	(0.184)	(0.192)	(0.190)	(0.182)	(0.081)	(0.062)
Mean Girls	2.503	2.661	2.743	3.161	3.596	6.771	3.560
Total obs.	3,791	4,799	5,067	5,236	5,317	5,617	5,617
Treatment – Boys	0.425**	0.484**	0.455**	0.431**	0.340**	0.092	0.000
	(0.179)	(0.199)	(0.192)	(0.176)	(0.165)	(0.101)	(0.057)
Mean Boys	3.584	3.920	4.221	4.776	5.259	7.986	5.693
Total obs.	4,702	5,570	5,783	5,913	5,969	6,258	6,258
p-value Girls=Boys	0.202	0.131	0.290	0.290	0.501	0.804	0.999

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A42: ATE on being in data (RCT analysis, bottom 85% sample) *without controls*, by gender.

	(1)	(2)	(3)	(4)	(5)
Panel A. Heterogeneity by gender, in a given year					
	Year 1	Year 2	Year 3	Year 4	Year 5
Treatment – Girls	0.022	0.008	0.006	0.009	0.007
	(0.021)	(0.015)	(0.019)	(0.018)	(0.018)
Mean Girls	0.667	0.787	0.752	0.762	0.751
Total obs.	5,633	5,633	5,633	5,633	5,633
Treatment – Boys	0.039**	0.032***	0.026*	0.033***	0.022*
	(0.015)	(0.011)	(0.013)	(0.012)	(0.013)
Mean Boys	0.739	0.825	0.800	0.814	0.805
Total obs.	6,283	6,283	6,283	6,283	6,283
p-value Girls=Boys	0.465	0.147	0.326	0.268	0.505
Panel B. Heterogeneity by gender, average across years					
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5
Treatment – Girls	0.022	0.006	0.003	0.003	-0.003
	(0.021)	(0.014)	(0.011)	(0.009)	(0.008)
Mean Girls	0.667	0.852	0.901	0.931	0.948
Total obs.	5,633	5,633	5,633	5,633	5,633
Treatment – Boys	0.039**	0.021**	0.019**	0.016**	0.012*
	(0.015)	(0.010)	(0.008)	(0.007)	(0.007)
Mean Boys	0.739	0.883	0.918	0.939	0.950
Total obs.	6,283	6,283	6,283	6,283	6,283
p-value Girls=Boys	0.465	0.313	0.169	0.195	0.094

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions are without controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is being in data in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is being in data in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A43: ATE on earnings (RCT analysis, bottom 85% sample) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment	187.200**	236.022**	309.593**	155.911	12.136	70.215	85.566
	(73.177)	(110.696)	(134.487)	(144.291)	(163.485)	(97.233)	(143.947)
Mean	1,222.664	1,698.941	2,070.003	3,197.317	4,136.427	7,224.677	10663.828
Total obs.	8,493	9,670	9,305	9,462	9,322	11,875	11,875
	Panel B. average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment	187.200**	206.926**	225.562**	213.203**	175.014	96.526	41.150
	(73.177)	(88.939)	(97.556)	(103.787)	(109.373)	(98.979)	(90.405)
Mean	1,222.664	1,398.823	1,571.501	1,926.608	2,295.399	5,232.391	4,132.031
Total obs.	8,493	10,369	10,850	11,149	11,286	11,875	11,875

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading. Mean is the average of the outcome variable in the control group. p<0.01, ** p<0.05, * p<0.10.

Table A44: ATE on earnings (RCT analysis, bottom 85% sample) *with controls*, by gender.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. Heterogeneity by gender, in a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Treatment – Girls	97.849	119.927	248.324	140.406	128.471	35.930	28.389
	(71.454)	(112.242)	(150.998)	(152.573)	(173.185)	(108.067)	(150.925)
Mean Girls	861.307	1,200.774	1,385.953	2,242.229	3,053.223	6,309.859	9,752.282
Total obs.	3,791	4,438	4,240	4,296	4,234	5,617	5,617
Treatment – Boys	238.135**	316.212**	342.819**	153.846	-98.365	88.391	121.085
	(99.201)	(150.736)	(163.687)	(186.758)	(204.969)	(159.568)	(202.337)
Mean Boys	1,523.854	2,140.556	2,667.527	4,027.049	5,074.548	8,064.007	11493.014
Total obs.	4,702	5,232	5,065	5,166	5,088	6,258	6,258
p-value Girls=Boys	0.163	0.211	0.582	0.945	0.261	0.736	0.692
	Panel B. Heterogeneity by gender, average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Treatment – Girls	97.849	97.091	141.428	135.784	131.476	74.440	24.324
	(71.454)	(85.913)	(97.618)	(101.327)	(106.083)	(97.539)	(108.156)
Mean Girls	861.307	987.434	1,078.357	1,331.380	1,618.225	4,206.908	2,469.996
Total obs.	3,791	4,799	5,067	5,236	5,317	5,617	5,617
Treatment – Boys	238.135**	283.930**	283.474**	265.791*	198.105	104.319	46.649
	(99.201)	(122.354)	(130.978)	(138.526)	(143.549)	(171.117)	(117.880)
Mean Boys	1,523.854	1,765.757	2,019.785	2,472.901	2,921.025	6,177.692	5,658.525
Total obs.	4,702	5,570	5,783	5,913	5,969	6,258	6,258
p-value Girls=Boys	0.163	0.132	0.283	0.341	0.625	0.838	0.872

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the months of employment in the year shown in the column heading for the subsamples of female and male, respectively. In Panel B the dependent variable is the average yearly months of work in the interval of years shown in the column heading for the subsamples of female and male, respectively. Mean is the average of the outcome variable in the control group of the indicated subsample. p<0.01, ** p<0.05, * p<0.10.

Table A45: ATE on taking the PSU entrance exam by year (RCT, bottom 85%)

	PSU 2018	PSU 2019	PSU 2020	PSU 2021	PSU 2022	PSU 2023
Treatment	-0.017	-0.002	0.001	0.000	0.001	0.001
	(0.027)	(0.012)	(0.005)	(0.003)	(0.002)	(0.002)
Total obs.	11916	11916	11916	11916	11916	11916
Mean	0.764	0.096	0.034	0.017	0.012	0.010

Notes: In this table we report the estimate for coefficient β in regression equation (1), estimated on the sample of students in the PACE experiment whose baseline grades place them in the bottom 85% of their school. This cohort graduated from high school in 2017. Standard errors clustered at the school level are shown in parentheses. *Mean* is the average of the outcome variable in the control group. Standard set of controls included. *Treatment* is a dummy variable equal to 1 if the student is in a school randomly allocated to PACE, to 0 otherwise.

Table A46: ATE on months of employment including labor market fixed effects (RCT, baseline top 85% *all* students) *with controls*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Main specification	0.330**	0.359*	0.457**	0.240	0.098	0.010	0.005
	(0.143)	(0.184)	(0.191)	(0.172)	(0.162)	(0.034)	(0.045)
Region FE	0.143	0.094	0.202	-0.039	-0.114	-0.021	0.024
	(0.125)	(0.139)	(0.167)	(0.149)	(0.149)	(0.035)	(0.048)
Province FE	-0.115	-0.082	0.084	-0.128	-0.085	0.006	0.061
	(0.106)	(0.131)	(0.154)	(0.147)	(0.152)	(0.036)	(0.049)
Municipality FE	-0.085	-0.052	-0.032	-0.250*	-0.047	0.041	0.096
	(0.103)	(0.153)	(0.152)	(0.143)	(0.134)	(0.044)	(0.071)
Observations	8493	9670	9305	9462	9322	11875	11875
	Panel B. Average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Main specification	0.330**	0.344**	0.359**	0.342**	0.292**	0.084	0.002
	(0.143)	(0.158)	(0.158)	(0.153)	(0.147)	(0.078)	(0.049)
Region FE	0.143	0.110	0.123	0.101	0.051	-0.039	0.021
	(0.125)	(0.121)	(0.121)	(0.119)	(0.115)	(0.067)	(0.042)
Province FE	-0.115	-0.088	-0.043	-0.044	-0.056	-0.036	0.052
	(0.106)	(0.108)	(0.108)	(0.111)	(0.109)	(0.071)	(0.040)
Municipality FE	-0.085	-0.066	-0.062	-0.068	-0.063	-0.004	0.030
	(0.103)	(0.117)	(0.106)	(0.106)	(0.101)	(0.094)	(0.060)
Observations	8493	10369	10850	11149	11286	11875	11875

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The first row corresponds to the main specification, the second to fifth rows add fixed effects for the local area, defined as: the region, the province, and the municipality. All regressions use the standard set of controls. In columns 1 to 5, the standard errors, reported in parentheses, are clustered at high school level. In columns 6-7, the standard errors, reported in parentheses, are bootstrapped using 100 replications and resampling at school level. In Panel A the dependent variable is the months of work in the year shown in the column heading. In Panel B, the dependent variable is the average yearly months of work across the years in the interval shown in the column heading. p<0.01, ** p<0.05, * p<0.10.

Table A47: ATE on earnings including labor market fixed effects (RCT, baseline top 85% *all* students) *with controls*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Panel A. In a given year						
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
Main specification	187.200**	236.022**	309.593**	155.911	12.136	70.215	85.566
	(73.177)	(110.696)	(134.487)	(144.291)	(163.485)	(97.233)	(143.947)
Region FE	87.770	78.432	117.807	-55.358	-163.691	-21.298	63.673
	(64.890)	(81.592)	(111.983)	(125.637)	(145.059)	(101.031)	(151.471)
Province FE	2.664	31.624	83.138	-45.498	-57.862	91.030	225.059
	(57.805)	(75.554)	(98.050)	(112.181)	(126.935)	(102.725)	(155.678)
Municipality FE	24.753	20.857	-11.104	-129.164	-59.510	121.738	264.597
	(62.463)	(79.502)	(92.240)	(109.589)	(131.919)	(134.757)	(221.865)
Observations	8493	9670	9305	9462	9322	11875	11875
	Panel B. Average across years						
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
Main specification	187.200**	206.926**	225.562**	213.203**	175.014	96.526	41.150
	(73.177)	(88.939)	(97.556)	(103.787)	(109.373)	(98.979)	(90.405)
Region FE	87.770	76.303	78.834	55.395	10.878	-59.390	26.999
	(64.890)	(66.576)	(73.562)	(80.682)	(87.424)	(93.070)	(93.591)
Province FE	2.664	21.267	31.011	20.797	4.250	22.848	127.626
	(57.805)	(60.082)	(63.661)	(70.283)	(75.845)	(94.410)	(92.264)
Municipality FE	24.753	19.566	3.124	-11.792	-16.595	46.712	111.547
	(62.463)	(61.676)	(58.606)	(62.493)	(69.477)	(123.392)	(134.901)
Observations	8493	10369	10850	11149	11286	11875	11875

Notes: This table shows the estimate of parameter β in equation (1), estimated in the sample of students who belong to the bottom 85% of their school based on baseline GPA. The first row corresponds to the main specification, the second to fifth rows add fixed effects for the local area, defined as: the region, the province, and the municipality. All regressions use the standard set of controls. In columns 1 to 5, the standard errors, reported in parentheses, are clustered at high school level. In columns 6-7, the standard errors, reported in parentheses, are bootstrapped using 100 replications and resampling at school level. In Panel A the dependent variable is yearly earnings in the year shown in the column heading. In Panel B the dependent variable is the average yearly earnings across the years in the interval shown in the column heading. Earnings are measured in thousands of December 2022 Chilean pesos. p<0.01, ** p<0.05, * p<0.10.

Table A48: LATE on earnings (RD) *with controls*, by perceived prob. of graduation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A. Heterogeneity by perceived prob. of graduation, in a given year							
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 10	Year 15
High perceived prob. of graduation							
Conventional	-73.788	533.420	493.551	47.134	-218.870	-321.723	-869.815
	(292.506)	(509.687)	(979.981)	(963.421)	(952.536)	(797.200)	(1304.347)
Robust	-50.964	662.768	694.739	314.267	-183.983	-497.805	-725.401
	(348.216)	(575.427)	(1198.404)	(1150.970)	(1140.005)	(936.498)	(1541.473)
Bandwidth	63.061	72.942	54.067	74.215	93.884	55.750	31.897
Bandwidth obs.	299	386	294	378	497	327	187
Mean	895.297	851.513	1,354.852	2,765.255	3,505.425	8,927.316	13449.789
Low perceived prob. of graduation							
Conventional	-870.896	-637.493	-88.832	546.825	-440.751	822.829	820.989
	(560.634)	(658.349)	(754.258)	(853.654)	(1023.398)	(833.400)	(1014.585)
Robust	-872.177	-484.010	-262.545	491.891	-651.829	1011.325	1039.155
	(616.065)	(767.135)	(901.588)	(1013.000)	(1141.804)	(936.944)	(1175.605)
Bandwidth	45.888	50.579	71.589	79.896	61.373	57.282	49.066
Bandwidth obs.	149	168	244	261	204	231	200
Mean	1,458.693	1,499.973	1,665.466	2,801.344	4,301.040	8,018.498	11712.087
p-value High=Low	0.180	0.081	0.602	0.725	0.871	0.241	0.268
Panel B. Heterogeneity by perceived prob. of graduation, average across years							
	Year 1	Years 1-2	Years 1-3	Years 1-4	Years 1-5	Years 6-10	Years 11-15
High perceived prob. of graduation							
Conventional	-73.788	326.829	356.828	283.157	211.216	-135.828	-451.203
	(292.506)	(400.758)	(591.510)	(638.801)	(668.163)	(783.537)	(464.498)
Robust	-50.964	398.938	420.734	408.754	345.898	-204.129	-606.129
	(348.216)	(453.192)	(717.206)	(763.318)	(788.011)	(906.216)	(521.436)
Bandwidth	63.061	69.328	56.322	57.163	60.392	61.552	45.430
Bandwidth obs.	299	382	323	330	349	357	267
Mean	895.297	843.348	1,060.044	1,513.181	1,888.843	5,484.839	5,827.081
Low perceived prob. of graduation							
Conventional	-870.896	-743.233	-572.888	-181.327	-448.672	500.502	648.936
	(560.634)	(545.921)	(507.457)	(577.482)	(618.930)	(877.639)	(645.061)
Robust	-872.177	-704.387	-621.548	-259.525	-518.233	462.160	864.928
	(616.065)	(640.795)	(611.497)	(673.093)	(715.149)	(969.872)	(731.995)
Bandwidth	45.888	48.045	49.634	63.649	65.915	55.078	52.901
Bandwidth obs.	149	171	189	239	249	221	211
Mean	1,458.693	1,385.704	1,251.836	1,577.327	2,064.152	5,186.190	4,911.543
p-value High=Low	0.180	0.055	0.203	0.545	0.411	0.482	0.108

Notes: This table shows the estimate of parameter δ in equation (2). All regressions use the standard set of controls. In columns 1 to 5, the standard errors reported in parentheses are clustered at high school level. In columns 6-7, the standard errors reported in parantheses are bootstrapped using 100 replications and resampling at school level. All panels show the observed impacts in each of year 1-5 and the predicted impacts in year 10 and 15 after high school graduation. In Panel A the dependent variable is the earnings in the year shown in the column heading for the subsamples of above median and below median perceived prob. of graduation, respectively. In Panels B the dependent variable is the average yearly earnings in the interval of years shown in the column heading for the subsamples of above median and below median perceived prob. of graduation, respectively. Mean is the average of the outcome variable just below the cutoff. *Robust* uses the robust approach with bias-correction suggested by Calonico, Cattaneo, and Titiunik (2014). Optimal bandwidth (Calonico, Cattaneo, and Titiunik, 2014), a linear polynomial of the ranking score and uniform kernels are used in all the specifications. p<0.01, ** p<0.05, * p<0.10.

C Technical Appendix

C.1 Variable construction

Labor force participation. The number of months employed in a year is calculated by summing the number of months an individual is observed participating in the labor market according to unemployment insurance (SC) data. An individual is considered as working in a month if he or she is observed in the SC dataset that month. This variable is set to zero if an individual is not observed in the SC dataset but is observed participating in education. The variable is set to missing for students who are not observed in the SC nor in the higher education datasets. Average months employed in an interval of years from t to T is the average number of months employed in the non-missing years from t to T. It is missing for individuals with missing months employed in all years between t and T.

Earnings. As in other studies using the SC dataset (Neilson et al., 2021), we collapse earnings observations for an individual to the annual level. This variable is obtained by summing all earnings appearing in the SC dataset during the year. If a person is observed in the education dataset in a specific year and no earnings are reported in the SC dataset, the yearly earnings are set to zero. If a person is not observed in the education dataset in a specific year and is not observed in the SC dataset at any point during the year, earnings are set to missing in that given year. Earnings are expressed in $1,000$ December 2022 Chilean pesos (exchange rate: CLP $1,000,000 \sim$ USD $1,000$). Average earnings in an interval of years from t to T is the average of earnings in the non-missing years from t to T. It is missing for individuals with missing earnings in all years between t and T.

Sectors. To construct the surrogate index, we use information on the sector of employment in each year. Individuals may work in more than one sector in a given year. In those cases, we choose the occupation that generated the highest earnings within the year. We use the following categorization of sectors:

- Agriculture, livestock, fishing
- Mining
- Manufacturing
- Electricity, gas
- Water
- Construction
- Sales and car reparation
- Transport
- Hospitality

- Communication
- Finance and banking
- Real estate
- Science
- Administrative services
- Public administration
- Education
- Health
- Arts and recreation
- Other services
- Home services
- Abroad
- Missing sector

Higher education outcomes. We consider any kind of higher education institution and, separately, selective colleges. Selective colleges are the 39 colleges that participated in the centralized admission system in the study period. The category 'any institution' comprises, beyond selective colleges, non-selective colleges and vocational institutes. Non-selective colleges offer academic programs like selective colleges, but did not participate in the centralized admission system in the study period (Ministry of Education, 2023). Vocational institutes comprise Technical Training Centres (*Centros de Formación Técnica*) and Technico-Vocational Institutes (*Instituto Técnico Profesional*), offering shorter degree programs (2-3 years vs 5 and more years) of a vocational nature. Like non-selective colleges, vocational institutes do not participate in the centralized admission system.

Let the institution index k refer to the type of higher education institution (selective college or any kind), and the time index t refer to the years after high school. The outcome variables are constructed as follows:

- *Enrolled in higher education institution of type k in year t:* equal to 1 if the student was enrolled in a higher education institution of type k in year t, and equal to 0 otherwise.
- *Graduated from higher education institution of type k by year t:* equal to 1 if a student graduated from an institution of type k in year t or before, or if the student is enrolled in an institution of type k in year t (and, therefore, can potentially graduate), and equal to 0 otherwise.
- *Ever enrolled in higher education institution type k by year 6:* equal to 1 if a student has been enrolled in a higher education institution of type k in any year from 1 to 6, and equal to 0 otherwise.

- *Dropout from higher education institution type k in year 6:* equal to 1 if a student enrolled in a higher education institution of type k in a year before 6, and in year 6 the student is not enrolled and has not graduated from higher education, and equal to 0 otherwise.
- *Selectivity:* Define a program as a major and institution pair. For the first program a student ever enrolls in, this variable is equal to the average SIMCE test score of all students who enrolled in the program (through the non-preferential admission channel) in the same year.
- *Ability distance:* For each program and student pair, this variable is equal to the difference between the program selectivity and the SIMCE test score of the student. It is calculated for the first program a student ever enrolls in.
- *Rank:* For each program and student pair, this variable is equal to the positional rank of the student in the program in terms of SIMCE test score. A value of 1 indicates that the student is the top-ranking one, a value of 0 indicates that the student is the bottom-ranking one. It is calculated for the first program a student ever enrolls in.

Expectations. In 2017 we administered a survey to students in the experimental cohort, during the last weeks of 12^{th} grade. In this study, we use answers to a question regarding the likelihood of graduating from a selective college conditional on enrolling in one. The English translation reads: "If I enroll in a university (not a Technical Training Center or Professional Institute) thanks to a high PSU score, I will complete my studies". The possible answers were: "Completely certain that I will not", "More likely that I will not", "Equally likely that I will and will not", "More likely that I will", "Completely certain that I will".[39] Extensive focus groups with students indicated that adding the wording 'thanks to a high PSU score' was necessary to ensure students understood the question was about selective colleges, which require taking the PSU entrance exam and obtaining a score above an admission cutoff. Given that the vast majority of students in our sample are confident their PSU score *will* be high enough to obtain an admission (see the detailed analysis of pre-college beliefs in Tincani, Kosse, and Miglino (2024)), we are confident students interpreted this question how we intended: "If I enter a selective college, I will graduate."

To build the variable *overconfidence*, we compare each student's answer to this survey question to her true probability of completing her studies conditional on enrolling in a selective college. We predict each student's true probability of graduating from a selec-

[39]The original Spanish text is: *Si ingreso a una universidad (no a un Centro de Formación Técnica o Instituto Profesional) por un alto puntaje PSU, terminaré mis estudios.* The possible answers were: *Totalmente seguro que no, Es más probable que no, Es igualmente probable que sí y que no, Es más probable que sí, Totalmente seguro que sí.*

tive college using a LASSO Probit regression estimated on the sample of students who have enrolled in selective college, using potential graduation from selective college as the dependent variable. The LASSO regression selects the best predictors among the set of standard controls, their second and third power, and their interactions. We then apply the estimated LASSO model to the sample of all survey respondents. Overconfidence is measured by the difference between the perceived and the actual probability of graduating from a selective college. The perceived probability is obtained by assigning numerical values to the survey answers. For robustness, we examine sensitivity to alternative numerical assignments

C.2 Surrogate index

To predict long-term effects on labor market outcomes we apply the surrogate index method described in Athey, Chetty, Imbens, and Kang (2019). We use data from five older cohorts of students to combine several shorter-term outcomes, such as educational performance and earnings up to five years after high school, into a surrogate index, i.e., the predicted value of the longer-term outcome. Table A2 provides summary statistics for this sample of older cohorts.

To predict each long-term outcome, such as, for example, earnings fifteen years post high school, we estimate the following regression model using data on the older student cohorts:

$$Y_{ist} = \beta_{0t} + \beta_{1t} Z_i + \beta_{2t} X_{is} + u_{it} \tag{4}$$

where Y_{ist} is the outcome t years post high school for student i, who attended school s, Z_i is a vector of intermediate education and labor market outcomes, and X_{is} is a vector of time-invariant student and high school characteristics, and characteristics of the labor market at the time of high school graduation. The parameters are indexed by t because we estimate a separate regression for each year post high school.

Vector Z_i includes intermediate outcomes pertaining education and labor force participation, each measured yearly during the first five years post high school, $t = 1, ..., 5$. The higher education outcomes include:

- enrollment in any type of higher education institution,
- enrollment in a selective college,
- graduation from any institution,
- graduation from selective college,
- Interactions of each of the above variables with the major area,
- Enrollment followed by dropout from any institution,
- Enrollment followed by dropout from a selective college.

The labor market outcomes include:

- yearly months employed,
- extensive margin labor force participation (LFP), categorizing individuals as working in the private sector, attending higher education, neither, or both,
- interactions: yearly months employed interacted with the occupational sector, LFP interacted with yearly months employed, and LFP interacted with yearly months employed and with the occupational sector,
- yearly earnings,
- interactions: yearly earnings interacted with the occupational sector, LFP interacted with yearly earnings, and LFP interacted with yearly earnings and with the occupational sector.

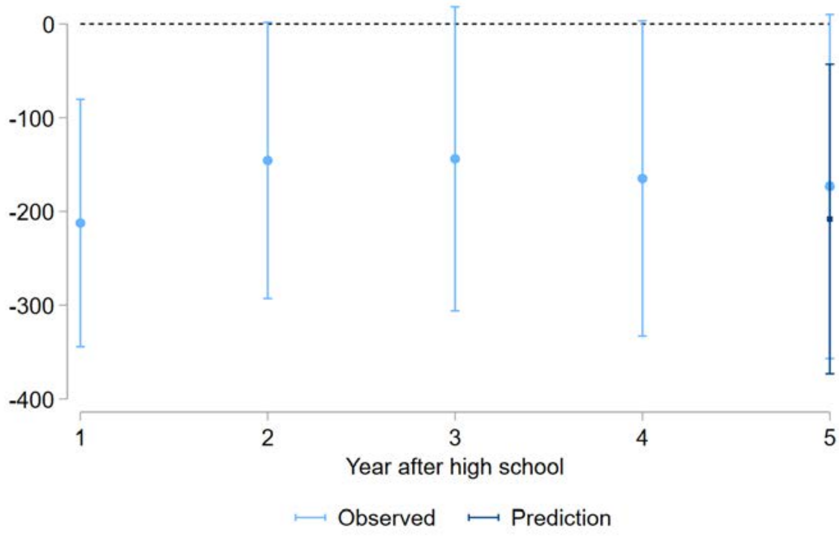
Vector X_{is} includes: student gender, age, indicator for having never failed a grade, high school track (academic or vocational), gender-specific national unemployment rate the year after high school graduation, high school GPA, indicator for whether the high school GPA is missing, school rurality, whether the school is in the Santiago metropolitan area, whether the school offers only academic tracks, cohort size.

Armed with the estimated coefficients from the regressions in (4), we use the shorter-term, intermediate outcomes to compute the value of the surrogate index for our study samples of younger students, $\hat{Y}_{ist}$. We then estimate the treatment impacts on this index, estimating equations (1) and (2) using $\hat{Y}_{ist}$ as dependent variables, and bootstrapping the standard errors to account for the estimation error in the surrogate index.

The average treatment effect on the surrogate index equals the treatment effect on the yet-unobserved future long-term outcomes under the assumptions that: 1) the treatment is orthogonal to potential outcomes in the main samples (RCT and RDD)[40]; 2) the conditional distribution of the long-term outcomes given the surrogate index is the same in the prediction and main samples; 3) the long-term outcomes are independent of the treatment conditional on the surrogate index. Implementing the technique proposed in Athey et al. (2019), we validate the third assumption by comparing the impact estimates on earnings and months worked at five years post high school, which are observed, to those obtained using the surrogate index for year five earnings and for year five months worked calculated using data from the first three years after high school. The results of the validation are reported in Figures A13 to A16, which show a close alignment between actual and predicted effects. Finally, Figures A17 to A20 show that there is also a close alignment between predicted and observed outcomes.

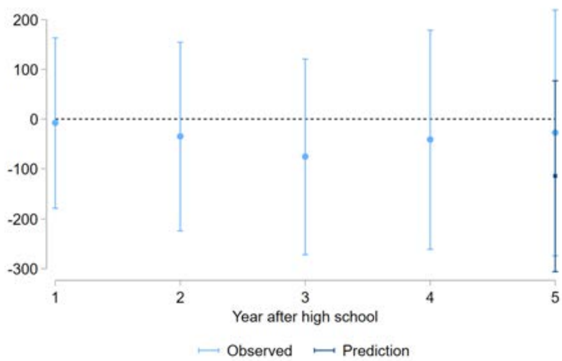
[40]This assumption can be applied to the RDD setting under a local randomization assumption: the treatment is 'as good as random' in a small window around the cutoff.

Figure A13: Validation of predictions on average yearly earnings, RD sample

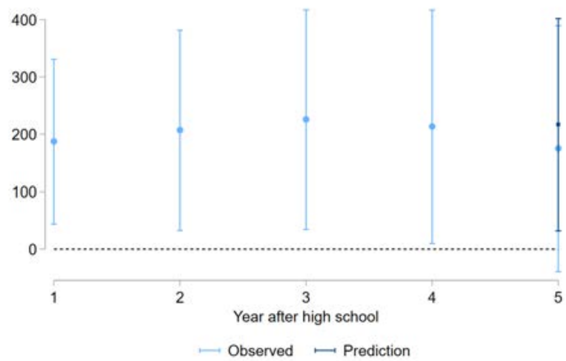


Notes: This figure shows the estimate of parameter δ in equation (2) in the RDD sample of students around the cutoff of 15% of the PRN score (based on GPA over the four years of high school). *Observed* are the point estimates when the dependent variable is the observed average yearly earnings in each of the first five years after high school. *Prediction* are the point estimates when the dependent variable is the predicted average yearly months of employment in year 5 using observed data up to year 3 after high school for the prediction. The standard errors are clustered at school level. The controls include: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, and type of high school track (academic or vocational).

Figure A14: Validation of predictions on average yearly earnings, experimental sample



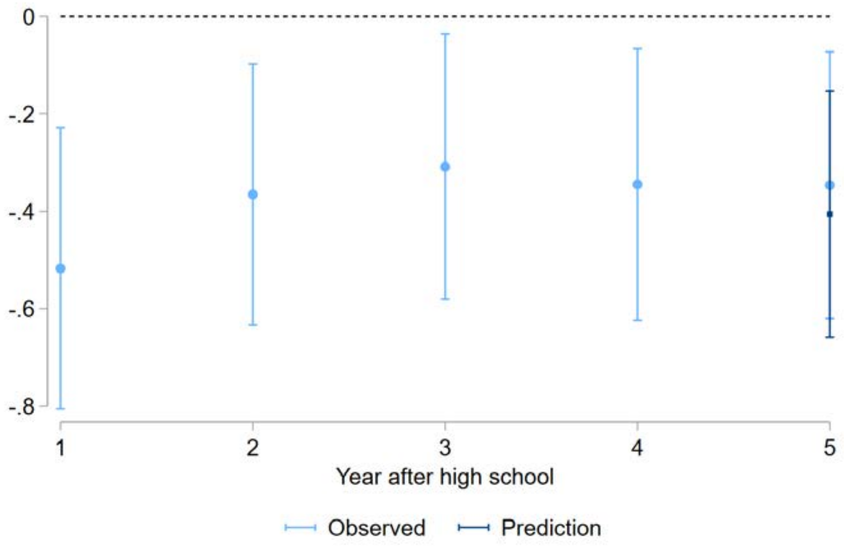
(a) Top 15%



(b) Bottom 85%

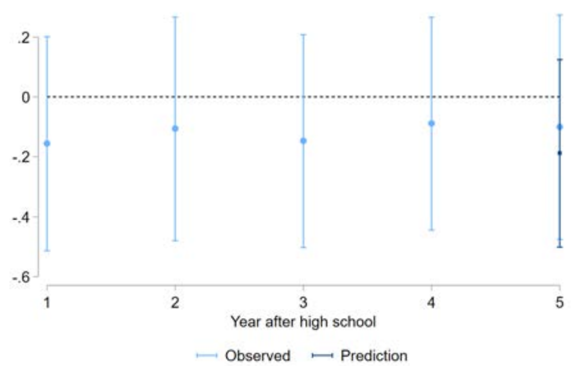
Notes: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the top 15% and the bottom 85% of their baseline GPA in their school. *Observed* are the point estimates when the dependent variable is the observed average yearly earnings in each of the first five years after high school. *Prediction* are the point estimates when the dependent variable is the predicted average yearly months of employment in year 5 using observed data up to year 3 after high school for the prediction. The standard errors are clustered at school level. The controls include: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, and type of high school track (academic or vocational).

Figure A15: Validation of predictions on average yearly months of employment, RD sample

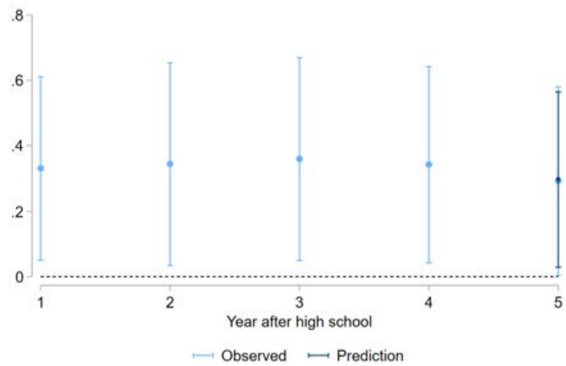


Notes: This figure shows the estimate of parameter δ in equation (2) in the RDD sample of students around the cutoff of 15% of the PRN score (based on GPA over the four years of high school). *Observed* are the point estimates when the dependent variable is the observed average yearly months of employment in each of the first five years after high school. *Prediction* are the point estimates when the dependent variable is the predicted average yearly months of employment in year 5 using observed data up to year 3 after high school for the prediction. The standard errors are clustered at school level. The controls include: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, and type of high school track (academic or vocational).

Figure A16: Validation of predictions on average yearly months of employment, experimental sample



(a) Top 15%



(b) Bottom 85%

Notes: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the top 15% and the bottom 85% of their baseline GPA in their school. *Observed* are the point estimates when the dependent variable is the observed average yearly months of employment in each of the first five years after high school. *Prediction* are the point estimates when the dependent variable is the predicted average yearly months of employment in year 5 using observed data up to year 3 after high school for the prediction. The standard errors are clustered at school level. The controls include: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, and type of high school track (academic or vocational).

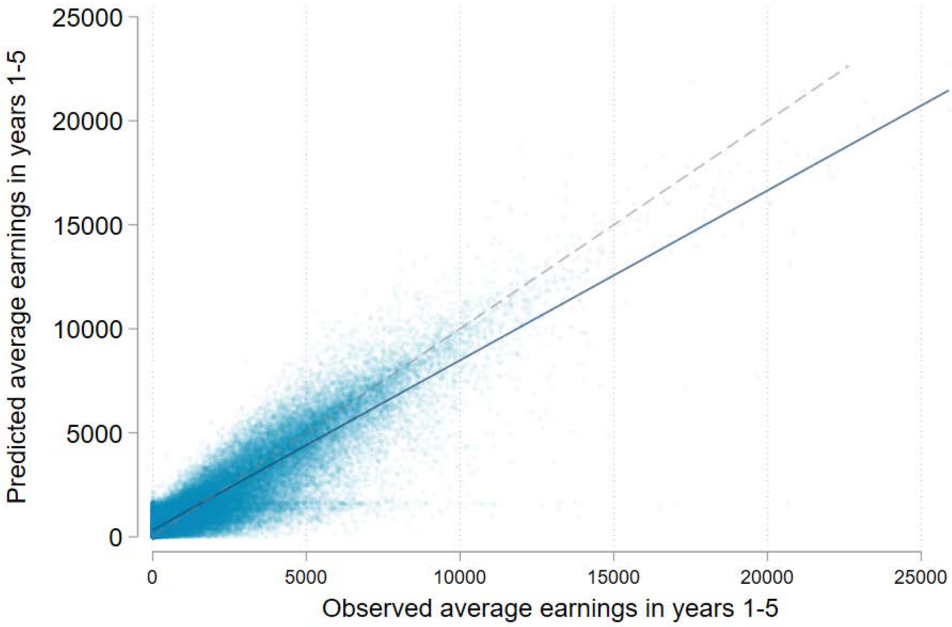
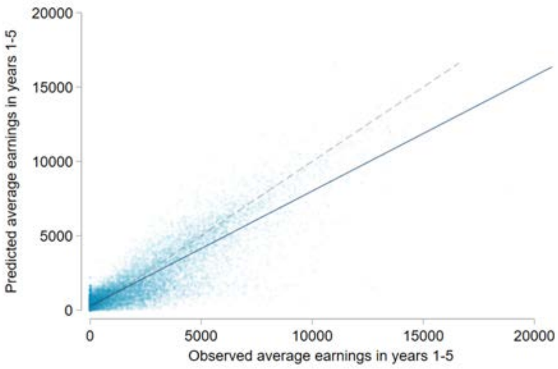
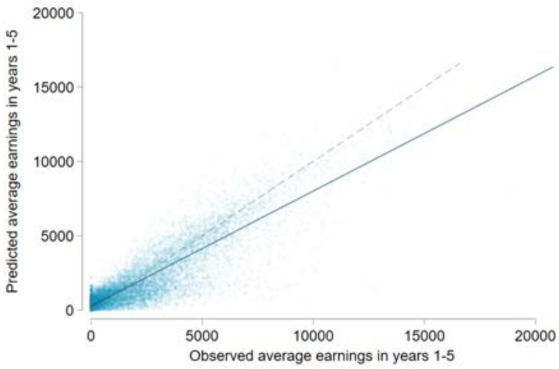


Figure A17: This figure plots observed and predicted cumulative earnings by the fifth year after high school for the RD sample. Predicted earnings are calculated as a surrogate index, a function of baseline characteristics and outcomes up until the third year after high school, estimated from the older cohorts of students and applied to the study population. Earnings are expressed in thousands of December 2022 Chilean pesos. The dashed line is the 45 degree line, the solid line is a linear fit.



(a) Top 15%



(b) Bottom 85%

Figure A18: This figure plots observed and predicted cumulative earnings by the fifth year after high school for top 15% and bottom 85% students. Predicted earnings are calculated as a surrogate index, a function of baseline characteristics and outcomes up until the third year after high school, estimated from the older cohorts of students and applied to the study population. Earnings are expressed in thousands of December 2022 Chilean pesos. The dashed line is the 45 degree line, the solid line is a linear fit.

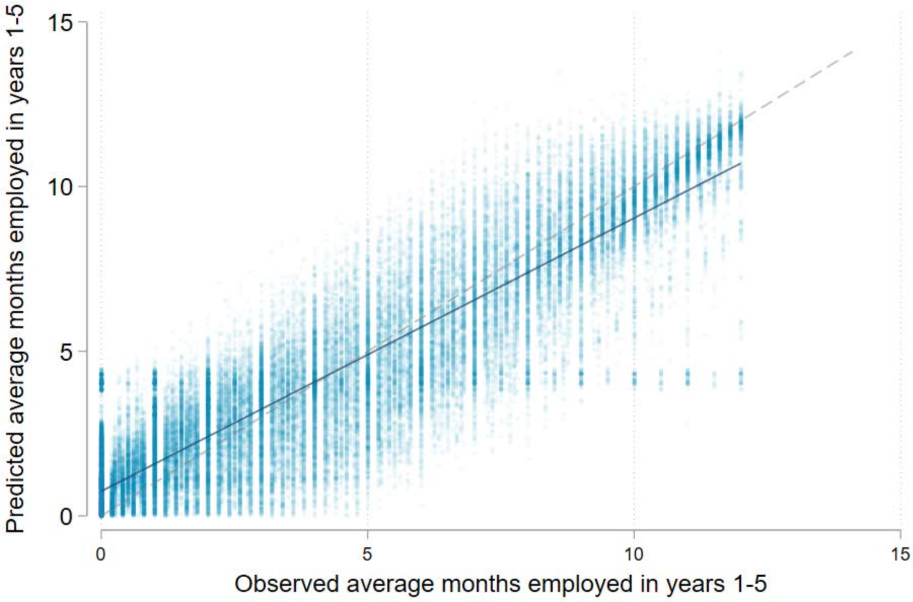
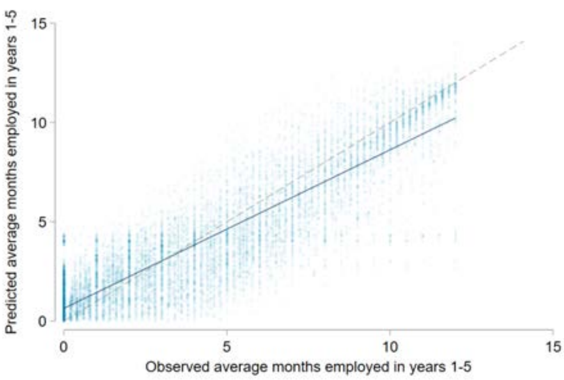
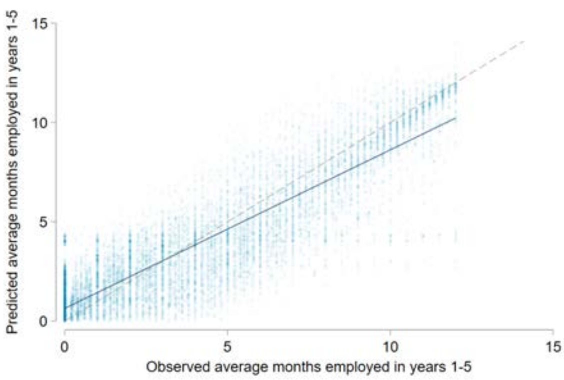


Figure A19: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the top 15% (a) and the experimental sample of students who belong to the bottom 85% (b) of their baseline GPA in their school. Predicted months of employment are calculated as a surrogate index, a function of baseline characteristics and outcomes up until the third year after high school, estimated from the older cohorts of students and applied to the study population. The dashed line is the 45 degree line, the solid line is a linear fit.



(a) Top 15%



(b) Bottom 85%

Figure A20: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the top 15% (a) and the experimental sample of students who belong to the bottom 85% (b) of their baseline GPA in their school. Predicted months of employment are calculated as a surrogate index, a function of baseline characteristics and outcomes up until the third year after high school, estimated from the older cohorts of students and applied to the study population. The dashed line is the 45 degree line, the solid line is a linear fit.

C.3 Estimation of bounds for RD estimates

In estimating impacts on the selectivity of the attended programs, we must account for sample selection due to any enrollment increase induced by PACE. We use Lee (2009) bounds in the experimental design. To obtain bounds on the RD estimates, we modify this procedure similarly to Dube, Giuliano, and Leonard (2019) and Dong (2019).

Let $r_{is} = p_{is} - c_s$ be the running variable centered at the cutoff. A PACE admission (defined as $A_i = 1$) leads to a Δ percentage point increase in the enrollment rate and $\tau = \frac{\Delta}{E(Enrollment_{is}|r^+_{is}=0, A_j=1)}$ is the ratio of the treatment effect on enrollment and the enrollment rate of the treated group at the cutoff.

Then, an upper bound for the RD estimate can be obtained by calculating the conditional mean $E(Y_{is}|r^+_{is} = 0, Y_{is} > Q_\tau, A_i = 1)$ for the right limit of the threshold where Q_τ is the τ^{th} quantile of outcome Y_{ij} at the right of cutoff, and comparing it to the unconditional mean $E(Y_{is}|r^-_{is} = 0, A_i = 0)$ for the left limit. For the lower bound, we need to estimate $E(Y_{is}|r^+_{is} = 0, Y_{is} < Q_{1-\tau}, A_i = 1)$ for the right limit of the threshold instead.

We implement this procedure by trimming the sample of observations on the right side of the cutoff within the bandwidth either from the top or the bottom of the outcome distribution at the appropriate quantile, and then performing the RD analysis on the restricted sample.

D PACE admission process

The PACE and regular applications must be submitted to the centralized system during the same time window. For the cohorts included in this study, the admission processes were separate. Students could submit a PACE application list with up to ten programs (i.e., university and major combinations), and, separately, a regular application list. Here we describe the PACE application and admission process. For each program listed in their PACE applications, applicants receive a distinct application score, called *Puntaje de Postulación PACE* (PPP). The score is calculated based on the applicant's GPA during the four years of high school and attendance during the 11^{th} and 12^{th} grades. To reduce the occurrence of identical scores across applicants, the score is adjusted for each program, taking into account the program's geographic location and its positional ranking within the applicant's list of preferences.[41]

Applicants to each program are ranked in descending order based on their application score, and available PACE slots are allocated according to this sequence. Should the number of applicants exceed the available slots, those not immediately admitted are placed on a waiting list for their first-choice program. Subsequently, these candidates

[41] The formula is $PPP = (0.8 * PRN + 0.2 * GPA) \cdot (1 + bonus_{attendance} + bonus_{geog}) + bonus_{listrank}$, where PRN is the *Puntaje Ranking de Notas* used to identify the top 15% of students, which is based on the high school GPA with some adjustments for the school context, and GPA is the raw high school GPA. The bonus for geographic location is awarded for applications to universities in the same area of Chile (North, Center, or South) as the student's high school, and the bonus for the rank of the program within the applicant list decreases with the program's rank.

are considered for admission to the programs listed as their second choice, following the same order-based allocation process. This procedure is iteratively applied to applicants' subsequent choices. Once an applicant is accepted into a program, they are automatically withdrawn from consideration for any programs ranked lower on their preference list. This measure ensures that no applicant is admitted to more than one program. However, applicants remain eligible for programs ranked higher than their successful application, should they be initially placed on a waiting list for such programs. In instances where a student eligible for a guaranteed PACE slot fails to secure admission in any of their listed preferences, the Ministry of Education employs a proprietary algorithm to determine their placement.