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

Michela Carlana†, Sara Chiuri‡, Enrico Miglino§, and Michela M. Tincani¶

July 2026

Abstract

Affirmative action could play a crucial role in fostering social mobility by bolstering the prospects of disadvantaged groups. This paper analyzes the long-term effects of a Chilean policy (PACE) that targets students in underprivileged schools, offering preferential admission to selective colleges to those graduating in the top 15% of their high school class. We leverage both the randomized expansion of the policy and the admission discontinuity. We find that PACE increases selective-college graduation without increasing dropout from higher education among top 15% students. It generates no significant average labor-market gains, reflecting offsetting gender patterns: women obtain selective-college degrees and more favorable longer-run income, while men have no selective-college degree gains and more unfavorable longer-run labor-market outcomes. For marginally eligible students, preferential admission increases selective-college graduation but also dropout from higher education, with no average labor-market gains. The dropout increase is concentrated among overconfident students, consistent with information frictions generating mismatch at the margin. Extrapolating away from the cutoff suggests that the access-mismatch tradeoff emerges around the 60^{th} percentile of national achievement. The results suggest a limit to how far preferential admission can go while delivering on its promise.

*We thank CEM (the research center of the Chilean Ministry of Education) and DIPRES (the budget office of the Chilean Ministry of Finance) for access to some of the data used in this study; Filippo Palomba, Liyang Sun, seminar participants at PUC, Tufts, LUISS, IFS, Queen Mary, UCL, Stockholm University, the VATT Institute, Rochester University, Cardiff Business School, University of Bologna, Tinbergen Institute, Bank of Portugal, CREST, Sciences Po, NHH, LSE, University of Surrey, University of Sussex, Yale, Carnegie Mellon University, and conference participants at NBER SI, CEP Education and Skills Conference, Stone Centre Conference on Economics and Inequality, Economics of Education Workshop at TSE, Opportunity Insights' Conference on Economic Mobility, Economics of Education Workshop at the University of Oslo, LEAP conference, and the Brucchi-Luchino Workshop on Labor Economics for helpful comments; and Sofía Betania Sierra Vásquez and Daniel Perez Huerta for excellent research assistance. AER RCT Registry ID: AEARCTR-0002288. The project obtained IRB approval from Harvard University (IRB23-0805) and UCL (10515/002), and is funded by the European Union (ERC, SOFIA, 101117537). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy.

†Bocconi University, NBER, CEPR, and LEAP-INSPIRE (e-mail: michela.carlana@unibocconi.it).

‡Princeton University and LEAP-INSPIRE (e-mail: sc1209@princeton.edu)

§Bank of Italy (e-mail: Enrico.Miglino@bancaditalia.it).

¶University College London, IFS, CEPR, CESifo, Stone Centre, and LEAP-INSPIRE (e-mail: m.tincani@ucl.ac.uk).

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 seek to reduce these gaps by improving admission chances for disadvantaged groups. Existing evidence shows that access to selective colleges can generate substantial benefits for disadvantaged students who are relatively well prepared (Bleemer, 2021; Black, Denning, and Rothstein, 2023). Less is known, however, about what happens when preferential admission reaches disadvantaged students with substantially lower academic preparation. This question has become increasingly important as institutions search for alternatives to race-based admissions and consider preferences based on socioeconomic disadvantage, which target students who often have fewer opportunities to build human capital before college (Cunha and Heckman, 2007) and lower achievement by the end of high school (Belley and Lochner, 2007; Bailey and Dynarski, 2011; Chetty et al., 2020; Hanushek et al., 2022). Whether preferential admission improves mobility for these students depends not only on expanding access, but also on whether students can convert access into degrees and later labor-market gains. A central question, therefore, is how far inclusion policy can expand while delivering on its promise of benefiting the students it targets.

We address this question in the context of a Chilean affirmative action policy called PACE (*Programa de Acompañamiento y Acceso Efectivo a la Educación Superior*), which targets disadvantaged high schools and offers preferential admission to selective colleges to students graduating in the top 15% of their high school cohort. Although top graduates within their schools, the high school standardized test scores of targeted students are 0.9 standard deviations (SD) below those of regular college entrants. For those just around the top 15% admission cutoff, this difference reaches 1.1 SD. This setting therefore provides a demanding test of how far preferential admission can be extended while still improving student outcomes.

Our empirical analysis combines two complementary research designs. First, we exploit the randomized expansion of the program in 2016 to estimate the effect of attending a PACE school for students in the top 15% of baseline high school GPA, the group targeted by the policy. The randomized controlled trial (RCT) estimand combines the effects of preferential admission with other PACE-school channels, including orientation activities and possible changes in incentives while in school (Tincani, Kosse, and Miglino, 2026). Second, we exploit the discontinuity in the admission rule to isolate the effect of receiving a preferential admission for students close to the eligibility cutoff. The regression discontinuity approach (RDD) focuses on students in schools participating in the PACE program since 2014. The designs therefore differ in both treatment and population: the

RCT estimates the effects of affirmative-action exposure for targeted students, while the RDD estimates the effects of preferential admission at the eligibility margin.

We evaluate impacts on students' educational attainment and labor-market outcomes, exploiting administrative records on education and formal-sector labor-market outcomes observed up to seven years after high school graduation, and we predict labor-market outcomes up to fifteen years using a surrogate-index approach (Athey et al., 2025). We also use survey data collected during students' final year of high school to measure beliefs about selective-college completion, with all sources linked by student identifiers.

Preferential admission can affect students through two opposing channels. Admission to a selective college may generate a positive quality effect by giving disadvantaged students access to better-resourced institutions, stronger peers, more valuable programs, and networks that can raise graduation rates and future earnings. But selective colleges may be more academically demanding. A long-standing theory, the 'mismatch hypothesis' (Sander, 2004), posits that preferential admissions targeted at low-achieving students could make them ex-ante 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.

Among students targeted by the policy, attending a PACE school increased selective-college enrollment by 29% and selective-college graduation by 24% relative to the control-group mean. Crucially, it did not increase dropout from higher education overall. Instead, the main additional margin was an increase in continued enrollment without a degree by year 7. Treated students worked and earned less in the short run, consistent with lower labor supply while enrolled in more demanding colleges. On average, these education effects did not translate into statistically significant gains in formal-sector earnings or monthly income at the observed labor market entry and the longer horizons we predict.

These average effects mask substantial gender heterogeneity. The gains in selective-college graduation are concentrated among women, who obtain additional degrees in both STEM and non-STEM fields and who, we predict, will experience longer-run income gains, potentially offsetting their short-term earning losses over their lifecycle. Men experience a less favorable pattern: PACE does not increase their selective-college graduation and lowers graduation from any higher education institution by year 7, likely reflecting a combination of delayed completion and possible degree losses, with less favorable predicted labor-market effects. Thus, PACE increases access for targeted students, but the degree and predicted-income gains are concentrated among women.

For students at the preferential-admission margin, the RDD estimates reveal a tradeoff. PACE admission increased selective-college enrollment by 18.2 percentage points (p.p.) and selective-college graduation by 7.8 p.p., 45% and 47% increases relative to the means just below the cutoff. But it also increased dropout from any higher education institution by 9.0 percentage points, a 43% increase: about 1.2 additional higher education dropouts for each additional selective-college graduate. Average effects on formal-sector annual earnings and monthly income are not statistically significant. At the admission margin, therefore, preferential admission generates selective-college degrees for some students, but leaves others more likely to drop out from higher education by year 7.

Effects are also heterogeneous by gender at the admission margin. PACE admission increases marginal women's selective-college graduation, especially in Non-STEM fields, reduces enrollment without a degree by year 7, consistent with faster resolution of education trajectories, and raises predicted monthly income conditional on working. Marginal men experience no selective-college degree gains and lower graduation from any higher education institution by year 7, especially in STEM fields; unlike in the RCT, this pattern is more consistent with lost degrees than delayed completion, as there is no offsetting increase in continued enrollment without (yet) a degree. Their predicted long-run annual earnings also fall. Thus, at the eligibility margin, preferential admission increases access and generates degree and predicted-income gains for women, while for men the admission-induced increase in dropout comes with degree losses.

Survey data point to information frictions as one mechanism behind these differences. Students substantially overestimate their probability of graduating from a selective college, and overconfidence is higher among men. Moreover, the increase in dropout at the admission margin is concentrated among students with above-median overconfidence, who are also more likely to enter programs in which their own baseline academic preparation is far below that of their peers. This evidence is consistent with the 'mismatch hypothesis' that affirmative action can generate mismatch effects in the presence of information frictions (Arcidiacono et al., 2011), and supports interpreting at least part of the admission-induced dropout as mismatch: some students accept preferential-admission offers expecting to graduate, but enter programs where completion is unlikely for them.

Next, we ask where the access-mismatch tradeoff begins as preferential admission is extended across the achievement distribution. Following Angrist and Rokkanen (2015) and Cingano et al. (2025), we extrapolate RDD estimates away from the PACE cutoff by using predetermined covariates to predict potential outcomes away from the discontinuity. The extrapolation suggests that preferential-admission eligibility would continue to generate selective-college graduation gains beyond the current top 15% cutoff, but that increases in higher education dropout first become detectable for students below roughly the top 6% of PACE high schools, corresponding to approximately the 60^{th} percentile

in the nationwide grade-10 test score distribution. While this threshold should be interpreted as an approximate benchmark rather than an exact rule, it highlights that an access-mismatch tradeoff can emerge as preferential admission is extended farther down the academic achievement distribution.

Finally, we use the coexistence of experimental and cutoff-based variation in the same empirical setting to link the RCT and RDD-based estimands. Building on a potential-outcomes framework in the spirit of Heckman and Navarro-Lozano (2004), we decompose the RCT estimand into an admission-eligibility component, identified through extrapolated RDD effects for eligible students, and a residual component capturing PACE-school exposure absent admission eligibility. We find that the key long-term education effects of attending a PACE school are largely accounted for by the preferential-admission eligibility component.

The paper contributes to the literature on affirmative action and preferential admissions in higher education in three ways. First, it extends evidence on percent-plan policies to a population with substantially lower pre-college academic preparation than those studied in prior work (Black, Denning, and Rothstein, 2023; Bleemer, 2021). In line with previous studies, we find positive effects on targeted students' selective-college enrollment and graduation (Arcidiacono and Lovenheim, 2016; Bleemer, 2022) and more favorable effects for women, as in Black, Denning, and Rothstein (2023), though in our setting graduation and predicted-income gains are fully concentrated among women. In contrast to previous affirmative-action studies, we find that preferential admission raised higher education dropout among marginally eligible students, and by extrapolating RDD estimates, we provide the first evidence on where a policy-relevant access-mismatch tradeoff can emerge.[1]

Second, we provide a direct test of the long-standing 'mismatch hypothesis' (Sander, 2004; Arcidiacono and Lovenheim, 2016). Building on the result that decision-making frictions are necessary for mismatch (Arcidiacono et al., 2011), we use linked survey-administrative data to measure these frictions. We show that dropout increases appear alongside belief errors, consistently with mismatch. Our findings suggest that mismatch is not a global property of preferential admission; it is a local phenomenon that depends on which students are induced to enroll.[2]

Finally, we link two common estimands in the literature. Prior studies have examined the impacts of broad affirmative-action program exposure, which may reflect preferential admission as well as other program components such as counceling or course requirements, and pre-college behavioral responses (Hinrichs, 2012, 2014; Bleemer, 2022; Black, Den-

[1]See also Otero, Barahona, and Dobbin (2023), who extrapolate away from RDD cutoffs using a structural model, and Bleemer (2023), who uses a linear extrapolation.

[2]These results also connect the affirmative action literature to evidence that returns to college may be limited for students with lower academic preparation (Heckman, Humphries, and Veramendi, 2018).

ning, and Rothstein, 2023; Antman, Duncan, and Lovenheim, 2024; Akhtari, Bau, and Laliberté, 2024; Oliveira, Santos, and Severnini, 2024). Other studies separately estimate the effects of preferential admission itself (Bagde, Epple, and Taylor, 2016; Francis-Tan and Tannuri-Pianto, 2018; Bleemer, 2021; Otero, Barahona, and Dobbin, 2023; Machado, Reyes, and Riehl, 2025). The decomposition provides a transparent bridge between these two estimands and shows when experimental evidence can serve as a benchmark for assessing RDD extrapolations (Angrist and Rokkanen, 2015), in the spirit of LaLonde (1986).

2 Institutional background

2.1 Higher education in Chile

Chile's higher education system is divided into three types of institutions: selective colleges, non-selective colleges, and vocational institutions. In 2018, selective colleges accounted for 41 percent of higher education enrollment, non-selective colleges for 8 percent, and vocational institutions for the remaining 51 percent.

Selective colleges refer to the thirty-nine institutions that participate in the centralized admission system (SUA, *Sistema Único de Admisión*).[3] These institutions offer academically oriented four- to six-year degree programs, and graduation rates conditional on pre-college achievement are lower than in other institutions (Figure A1). Admission requires students to take a national standardized entrance exam, known as the PSU (*Prueba de Selección Universitaria*) until 2020, submit ranked applications through a centralized platform, and be assigned to programs through a centralized allocation mechanism. Despite this centralized process, access to selective colleges is highly unequal across socioeconomic background in line with most industrialized countries: students from higher-income families are substantially more likely to enroll in this sector than students from lower-income families (Figure A2).

2.2 The PACE program

PACE (*Programa de Acompañamiento y Acceso Efectivo a la Educación Superior*) is a preferential admission policy introduced to expand access to selective colleges among students from disadvantaged high schools. The program combines three components: preparation and orientation during the last two years of high school, reserved seats in participating colleges, and academic support during the first year of higher education for students who enroll through the program. It is implemented through agreements between

[3]The list of all selective universities is available in Table A1.

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

PACE targets high schools selected by the government based on a school-level vulnerability index (*Índice de Vulnerabilidad Escolar*). The program was first introduced in the 69 most disadvantaged Chilean high schools in 2014, and later expanded to progressively less disadvantaged schools, reaching 456 by 2016. The first participating schools were located in the poorest regions of south-central Chile and in the poorest neighborhoods of the southwestern Santiago metropolitan area (Figure 1). When a school first enters PACE, the program initially applies to the cohort in 11^{th} grade; in subsequent years, it also applies to 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, based on students' socioeconomic status.[4] The average family income of students in these schools, for example, was half that of the average Chilean student (Tincani, Kosse, and Miglino, 2026). Following advice from our research team and using a randomization code written by PNUD Chile (United Nations Development Program), the government randomly selected 64 of the 218 eligible schools to receive the PACE treatment. Figure A3 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 the beginning of the school year, students who were enrolled in high schools randomly selected for treatment were informed their school was joining PACE. The announcement was made after the school enrollment deadline, ruling out strategic selection into treated high schools (Tincani, Kosse, and Miglino, 2026).

Students in participating PACE high schools can apply to a selective college through the regular centralized admission system, like any other student in Chile. In addition, they become eligible for a PACE admission if they satisfy three conditions:

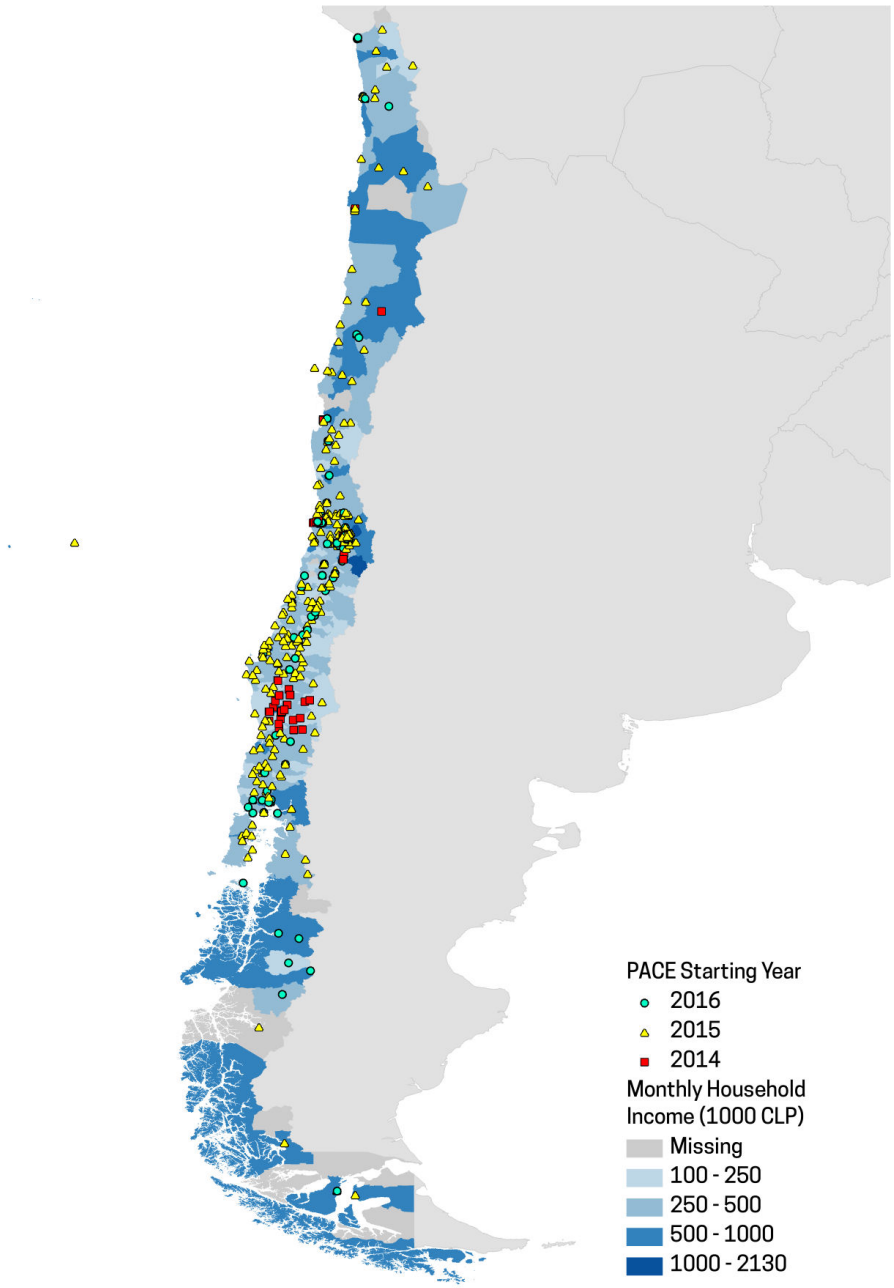
1. They rank above the 85^{th} percentile of the high school GPA distribution, or obtain a GPA score at or above a national cutoff.[5] This feature makes PACE a 'percent plan' program.

2. They attend a PACE school in grades 11 and 12 and complete the program's high school orientation activities.

[4]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.

[5]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 income of students' households (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.

3. They take the entrance exam for selective colleges, although the score does not affect eligibility for a PACE admission.

To receive a PACE admission, eligible students must apply through the PACE channel. During the cohorts we study, a PACE admission could be used only in the year immediately after graduating from high school, potentially inducing students to enter higher education directly after high school. Finally, PACE differs from other percent plans, such

as those implemented in Texas and California (Horn and Flores, 2003, 2015), as eligibility does not require additional coursework.

The selective college seats offered through PACE are roughly similar to those offered through the regular admission channel in terms of study field (STEM vs non-STEM), selectivity, and distance from the student's home (Tincani, Kosse, and Miglino, 2026). Seats offered through PACE were supernumerary, so they did not mechanically displace regular-admission seats. PACE is also small relative to the size of the higher education sector: in 2018, only 2.18% of all college enrollments occurred through PACE, distinguishing this program from large-scale affirmative action policies that induce large inflows of quota students into college (Machado, Reyes, and Riehl, 2025). This feature makes it unlikely that colleges adjusted grading or teaching standards in response to PACE. Thus, if PACE induces students to enroll in programs for which they are underprepared, we would expect this to appear primarily through persistence and dropout outcomes.

Finally, although PACE seats are not formally tied to financial aid, almost all students eligible for PACE seats are also eligible for tuition fee waivers, since students from families in the bottom 60 percent of the income distribution qualify for these waivers.[6]

3 Data

3.1 Data sources

We link individual-level administrative records from several sources, summarized in Table 1, through a collaboration with the Chilean Ministries of Education and Finance. The data cover the cohorts exposed to PACE between 2014 and 2016, when students were in 11th grade and PACE first enters their high schools. For the 2016 cohort, the data also include students in schools not exposed to the program as part of the experimental design described below. We also observe older cohorts of students who graduated from these same high schools between 2006 and 2010. The older cohorts allow us to relate early educational and labor-market outcomes to longer-run earnings, which we use to construct the surrogate-index measures described below.

First, we use school records from the Ministry of Education to obtain high school enrollment, grades, demographic characteristics, and school characteristics. Second, we use grade-10 standardized achievement test scores from SIMCE (*Sistema de Medición de la Calidad de la Educación*), which also include information from parent questionnaires, such as parental education and household income. Third, we use records from the SEP (*Subvención Escolar Preferencial*) program to identify students classified by the govern-

[6] 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.

Table 1: Overview of Data Sources

Dataset	Variables	Timing	Source
1. School records	High school enrollment, school characteristics, grades, demographic characteristics	Grades 9-12	Admin
2. *SIMCE*	Achievement test scores, background characteristics	Grade 10	Admin
3. *SEP*	Very-low-SES classification (*Prioritario* student)	Grade 10	Admin
4. Student survey	Expectations	Grade 12	Primary
5. Higher education records	PSU scores, engagement with centralized application system, enrollments, graduation and persistence in any higher education institution, program selectivity and major	Up to 7 years after high school	Admin
6. PACE program records	Allocation of PACE seats in selective colleges, enrollments, graduation and persistence via PACE channel, program selectivity and major, running variable (PRN)	Up to 7 years after high school graduation	Admin
7. Unemployment insurance	Months worked, earnings, occupational sector in any formal private-sector job	Up to 7 (actual) and 15 (predicted) years after high school	Admin

Notes : Table A2 reports which majors are categorized as STEM.

ment as very low socioeconomic status students. Fourth, 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.

Fifth, we use administrative records from the college entrance exam and centralized admissions system, which capture students' PSU scores and engagement with the regular centralized-admissions channel. Sixth, we use PACE program records, which identify students' Puntaje Ranking de Notas (the running variable for PACE eligibility), PACE admission offers, PACE seat allocation, and enrollment through the PACE channel. In both channels, we observe enrollment, persistence, dropout, and graduation at any higher education institution, through 2024, as well as program characteristics, including selectivity and major.

Finally, we link these education records to monthly formal-sector labor-market records from Chile's unemployment insurance system, the *Seguro de Cesantía*, housed at the Ministry of Finance. These records provide monthly taxable earnings, months worked, and sector of employment. The unemployment insurance data cover formal private-sector employment and therefore do not capture self-employment or public-sector employment. They also censor earnings above the taxable earnings cap, although this limitation is

unlikely to be important in our study population, where earnings are typically well below the cap (only 0.18% of observations in our sample are censored).

3.2 Samples and descriptive statistics

We use two main study samples. The first is the RCT top 15% sample, which includes students in the experimental schools whose baseline GPA placed them in the top 15 percent of their school. Baseline GPA is measured using grades 9 and 10, before students learned whether their school would participate in PACE. This sample captures the students targeted by the PACE admission rule and allows us to estimate the effect of attending a school assigned to PACE among high-achieving students in disadvantaged schools.

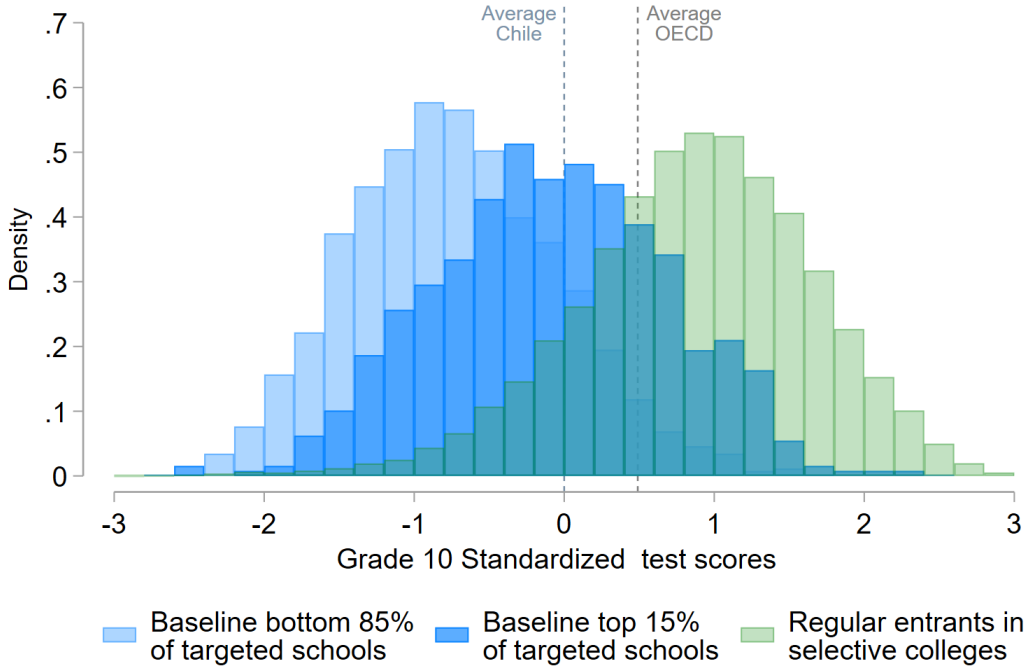
The second is the RDD sample, which includes students in PACE schools whose GPA ranking score placed them close to the preferential-admission cutoff. Unlike the RCT sample, which includes the cohort entering grade 11 in 2016, the RDD sample includes the cohorts entering grade 11 between 2014 and 2016. This sample allows us to isolate the effect of receiving a preferential admission offer among students close to the PACE eligibility threshold.

For the RCT sample, we define GPA rank using only grades 9 and 10 to avoid treatment-induced selection into the top 15 percent during grades 11 and 12. For the RDD sample, we use the policy-relevant GPA ranking score, the PRN, calculated from grades 9 through 12. This score is released to researchers only for students who registered to take the college entrance exam (PSU).

Students in our samples are academically disadvantaged relative to regular entrants to selective colleges. Regular entrants to selective colleges score, on average, at the 79^{th} percentile of the national grade-10 test-score distribution. By contrast, students in the top 15% of PACE schools score at the 52^{nd} percentile, and students close to the PACE admission cutoff score around the 46^{th} percentile. Figure 2 illustrates that the top 15% score 0.88 standard deviations below regular entrants in selective colleges. 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 0.20 standard deviations below the top 15% (Figure A4, Table 2).

PACE students are also less academically prepared than students targeted by comparable percent-plan policies in the United States. The top 15% RCT sample scores at approximately the 52^{nd} percentile of the national grade-10 test score distribution, compared with the 86^{th} percentile among students studied by Black, Denning, and Rothstein (2023) in their analysis of the Texas Top Ten Percent policy. Similarly, students in the PACE RDD sample score at approximately the 34^{th} percentile of the entrance exam distribution, compared with the 72^{nd} percentile among students studied by Bleemer (2021)

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 standardized 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 tenth 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 Tincani, Kosse, and Miglino (2026).

in his analysis of California's Eligibility in the Local Context policy.[7] The PACE RDD sample is closer to, although slightly below, the bottom quartile of the California sample. These comparisons underscore that PACE expands selective-college access to students with substantially lower measured academic preparation than those affected by the U.S. percent-plan policies most closely related to this study.[8]

Table 2 reports baseline characteristics for the two study samples. Female students are overrepresented among the highest-ranked students in PACE schools, especially near the preferential-admission cutoff. Around 60 percent of students in both samples are classified by the government as very low socioeconomic status, and roughly two thirds attended a vocational high school track. Students in the RDD sample are less academically prepared than students in the top 15% RCT sample, as measured both by grade-10 test scores and by grade repetition.

[7] Additionally, entrance-exam takers in the United States may be more positively selected than in Chile: while around 90% of the high school graduating cohort takes the exam in Chile, only around 60% do so in the United States.

[8] Appendix Table A5 reports the test-score summary statistics underlying these comparisons, including SIMCE and PSU percentiles for the PACE samples and the comparable samples from Black, Denning, and Rothstein (2023) and Bleemer (2021).

Table 2: Baseline characteristics of the RCT and RDD samples

	(1)	(2)	(3)	(4)	(5)	(6)
	Top 15%			RD		
	Obs.	Mean	S.D.	Obs.	Mean	S.D.
Female	2437	0.56	0.50	13048	0.59	0.49
Age	2437	16.32	0.64	13048	16.33	0.65
Very low SES	2437	0.59	0.49	13048	0.61	0.49
Mother's education	1914	9.68	3.10	8565	9.98	3.15
Father's education	1795	9.50	3.15	8090	9.79	3.24
Family income	1919	284.05	195.67	8554	310.16	255.54
SIMCE	2432	-0.01	0.82	12929	-0.21	0.76
Never failed	2437	0.94	0.24	13048	0.92	0.27
Santiago	2437	0.17	0.37	13048	0.22	0.41
Rural	2437	0.04	0.19	13048	0.05	0.22
Academic track	2437	0.31	0.46	13048	0.37	0.48

Notes: This table reports the number of observations, mean and standard deviations for pre-determined variables for the top 15% by baseline high school GPA ranking of the experimental sample, 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 higher education and labor market outcomes, referring the interested reader to Appendix B.1 for a detailed description of variable construction. For education, we measure whether, within seven years after high school graduation, a student ever enrolls in higher education (*Ever Enrolled*), obtains a degree (*Graduation*), drops out after enrolling (*Dropout*), or remains enrolled without a degree (*Still Enrolled*). We construct these outcomes separately for selective colleges, the direct target of PACE, and for any higher education institution, which also includes non-selective colleges and vocational institutions. This distinction allows us to study both direct effects on selective-college outcomes and substitution across sectors. Table A3 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. Around one in five students in the top 15% graduate from selective colleges within 7 years from high school. As expected, the probability of ever enrolling and graduation in higher education is a bit lower for students in the RDD sample.

For labor-market outcomes, we use annual formal private-sector earnings and employment from the unemployment insurance records. Observed outcomes cover the first seven years after high school graduation. We also predict medium- and longer-run outcomes using the surrogate-index method (Athey et al., 2025) described in Appendix B.2. This procedure uses older cohorts from the same high schools, for whom both early outcomes and longer-run labor-market outcomes are observed, to predict earnings and employment

for years 8 through 15 after high school.[9] We validate the surrogate-index procedure by using early outcomes to predict later outcomes that are already observed. Specifically, we predict earnings and monthly income in years 5–7 after high school using information available through year 4, and compare treatment-effect estimates based on these predictions with estimates based on the observed outcomes. Appendix B.2 shows that the predicted and observed treatment effects are closely aligned in both the RCT and RDD samples. This alignment likely reflects the richness of the prediction data and the small scale of PACE enrollment, which makes it unlikely that the program changed the value of college attendance through shifts in teaching standards or network opportunities, as was instead documented in larger programs such as the Brazilian quota system (Machado, Reyes, and Riehl, 2025).

In Table A3, we report the summary statistics of the average yearly earnings and number of months worked in the short term (1-7 years after high school graduation), in the medium term (8-11 years after high school graduation), and in the long term (12-15 years after high school graduation). In the short term, students in the top 15% of experimental schools worked for around 3.6 months and earned 2.1 million pesos, approximately 2,350 USD. Table A3 shows the expected life-cycle pattern in labor-market outcomes. Average annual earnings rise substantially over time in both study samples. In the top 15% RCT sample, earnings increase from about 2.1 million pesos (2,350 USD) in years 1-7 after high school graduation to 5.6 million pesos in years 8-11 and 6.4 million pesos in years 12-15. In the RDD sample, they increase from about 2.2 million pesos to 5.4 million pesos and 5.9 million pesos over the same horizons. Average months of formal employment also rise between the first seven years and years 8-11, but decline somewhat in years 12-15. This decline is driven by women, for whom formal employment falls at longer horizons.

4 Empirical methodologies

We use two complementary research designs. First, we implement a randomized controlled trial (RCT), exploiting the randomized expansion of PACE in 2016 to estimate the effect of attending a school assigned to the program. This design identifies the impact of exposure to PACE among students in the top 15% of the baseline GPA distribution, the group targeted by the policy. The estimated effect combines the offer of preferential admission for some with other components of the program, including preparation and orientation

[9] Appendix Table A4 provides a summary of the baseline characteristics and outcomes of the sample used to construct the predictions. As expected, these students display a 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.

during high school and any changes in incentives to study (Tincani, Kosse, and Miglino, 2026).

Second, we implement a regression discontinuity design (RDD), exploiting the discontinuity in the PACE admission rule at the top 15% GPA cutoff. This design compares students in PACE schools whose GPA ranking scores placed them just above and just below the preferential-admission threshold. It therefore isolates the local effect of receiving a PACE admission offer among students close to the eligibility cutoff.

The two designs answer related but distinct questions. The RCT estimates the effect of introducing PACE for high-achieving students in disadvantaged schools, using the cohort entering grade 11 in 2016. The RDD estimates the effect of preferential admission itself for marginally eligible students, using cohorts entering grade 11 between 2014 and 2016. Together, the two designs allow us to compare the effect of exposure to PACE for targeted students on average with the effect of extending preferential admission to students at the margin of eligibility. Section 6 returns to this comparison, decomposing the effect of attending a PACE school into preferential-admission eligibility and other PACE-school channels.

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:

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

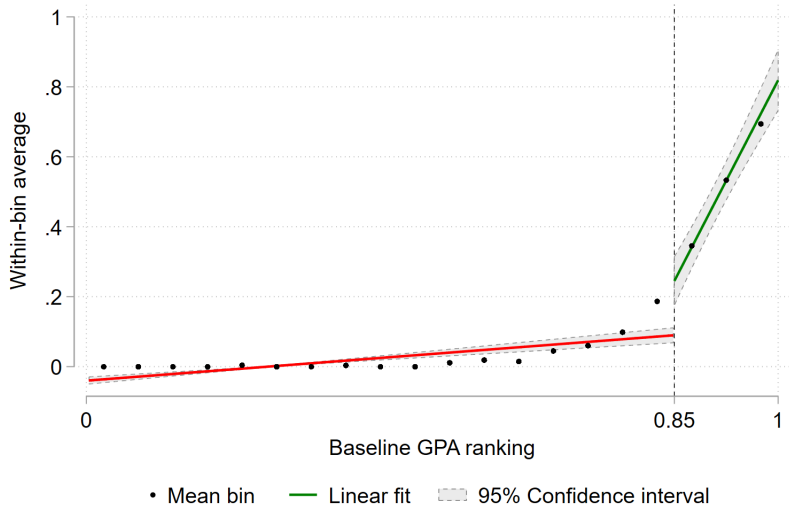
where Y_{is} is the outcome of student i in school s, T_s is a dummy variable equal to 1 if school s was randomly selected to participate in PACE and to 0 otherwise, and X_i is a vector of student i's baseline characteristics.[10] The standard errors are clustered at the high school level.

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% of GPA received a PACE admission offer (Figure 3).[11]

[10]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, the high school track (academic or vocational) and school-region fixed effects.

[11]The share of students receiving the offer is around 10.4% of the school population 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 attended the orientation classes) or they may not have applied.

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.

Table A6 shows that students in the top 15% RCT sample are balanced across treatment and control schools in predetermined characteristics, including gender, age, socioeconomic status, parental education, family income, baseline standardized test scores, grade repetition, residence in Santiago, school rurality, and high school track. All RCT specifications include the predetermined controls described above and region fixed effects; robustness specifications without controls are reported in the Appendix.[12]

4.2 Regression Discontinuity Design

The Regression Discontinuity Design (RDD) exploits the cutoff embedded in the PACE admission rule. Within each PACE school, students become eligible for preferential admission if their GPA ranking score places them above the top 15% cutoff, subject to the additional requirements described in Section 2.2. Crossing this cutoff generates a large discontinuity in the probability of receiving a PACE admission offer. Because not all students above the cutoff satisfy the remaining requirements or apply to college through the PACE channel, we implement a fuzzy regression discontinuity design (Lee and Lemieux,

[12]Because we also report heterogeneity by gender, Table A6 presents balance tests within this subgroup. Treatment and control students remain balanced within these subsamples. 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% at baseline is 16.98% because there are students with the same GPA average at baseline.

2010).[13] We estimate the equation:

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

where Y_{is} is the outcome for student i who attended high school s, A_i is equal to 1 if student i receives a PACE admission offer, and X_i is a vector of student characteristics.[14] The term η_s denotes high school fixed effects. The running variable p_{is} is the GPA score used to rank student (PRN) by the policy, and c_s is the school-specific cutoff for preferential admission, defined as the minimum between the 85^{th} percentile of the PRN score in school s and the national cutoff. We instrument A_i using the following first-stage regression:

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

The functions $f(p_{is} - c_s)$ and $g(p_{is} - c_s)$ are local linear functions of the running variable, normalized relative to the cutoff, and are allowed to differ 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. The standard errors are clustered at the high school level.

The parameter δ identifies the impact of receiving a PACE admission for the subset of compliers near the cutoffs. The exclusion restriction requires that crossing the threshold affects long-term outcomes exclusively through the PACE. Given that all students in the high school receive counseling before knowing whether they scored above the threshold, crossing the cutoff is likely to affect outcomes solely by triggering the PACE offer, including the preferential admission and tutoring, rather than by altering students' information.

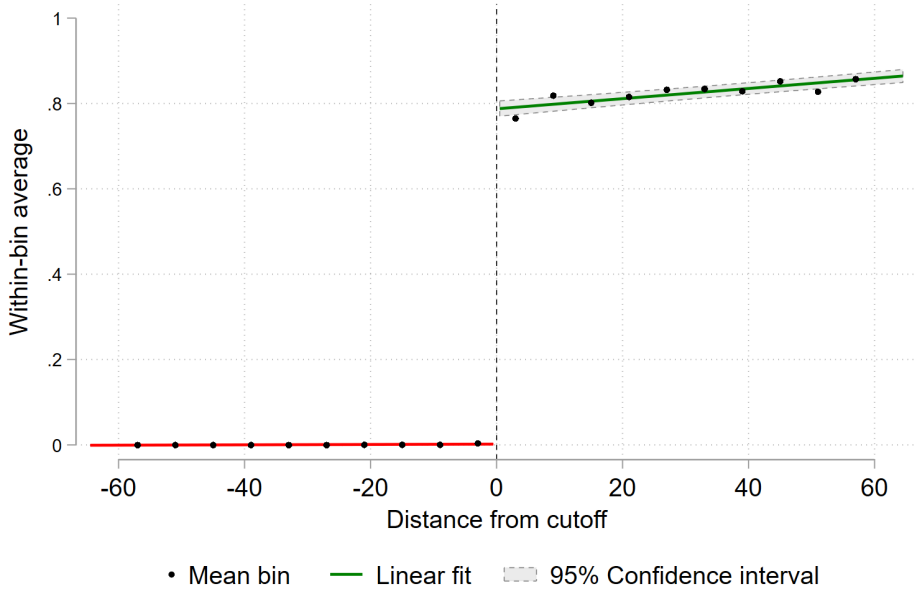
For each outcome, we restrict the estimation sample to students within the optimal bandwidth around the cutoff computed according to Calonico, Cattaneo, and Titiunik (2014). We report robust estimates with bias-correction obtained following Calonico, Cattaneo, and Titiunik (2014) and Calonico et al.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

Figure 4 shows the first stage: the probability of receiving a PACE admission offer as a function of the distance between a student's PRN score and the preferential-admission cutoff. No student below the cutoff received a PACE admission offer, while 85.36 percent of students above the cutoff did, generating a clear discontinuity at the cutoff. The probability of receiving an offer increases slightly with the student's ranking score.

[13] The running variable is available to us only for students who took the entrance exam, therefore, this criterion is satisfied for all students in this analysis.

[14] 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).

Identification requires that students cannot precisely manipulate their position relative to the cutoff. This assumption would fail if, for example, students or schools could adjust grades to move students just above the threshold. We test for manipulation using the density of students around the cutoff, following McCrary (2008). The test does not reject the null hypothesis of no discontinuity in the density (p-value of 0.697, Figure A5). We also find no evidence of manipulation using the test proposed by Cattaneo, Jansson, and Ma (2022), which does not require pre-binning the data (p-value 0.940).

Identification also requires that students just above and just below the cutoff are comparable in predetermined characteristics (Lee and Lemieux, 2010). Table A7 reports discontinuity tests for baseline covariates, estimated using the first-stage specification with each predetermined characteristic as the dependent variable and without controls. Only one of the eleven estimates, for age, is statistically significant at conventional levels; students above the cutoff are less than six weeks older than students below it. The main specifications control for age and other predetermined characteristics. Overall, the evidence supports the interpretation that PACE admission offers are locally as good as randomly assigned around the cutoff.

5 The impacts of PACE on targeted students

5.1 Conceptual framework

Preferential admissions can affect targeted students through two opposing forces. On the one hand, admission to a selective college may generate a positive quality effect. Selective colleges typically offer greater resources, more academically prepared peers, stronger faculty, and access to programs and networks that may raise graduation rates and future earnings. On the other hand, selective colleges may also be more academically demanding. If students induced to enroll through preferential admission are not sufficiently prepared for these demands, they may be more likely to struggle, delay progress, or drop out. In that case, preferential admission could generate a negative match effect (Arcidiacono and Lovenheim, 2016).

The mismatch hypothesis emphasizes this second possibility (Sander, 2004).[15] For students who would otherwise enroll in less selective institutions, preferential admission may redirect them toward selective colleges where completion is less likely. For students who would otherwise not enter higher education immediately after high school, preferential admission may instead operate on the extensive margin, inducing enrollment that is followed by dropout. In both cases, the short-run costs of enrollment, which could include foregone earnings, may not be offset by future returns if students do not complete a degree, lowering their lifelong earnings.

Existing evidence on percent-plan policies has generally found positive effects of access to selective colleges for disadvantaged students (Black, Denning, and Rothstein, 2023; Bleemer, 2022, 2021). However, as discussed in Section 3.2, PACE reaches students with substantially lower measured academic preparation than those affected by the U.S. percent-plan policies most closely related to this study. Whether preferential admission remains beneficial for this population, and for which students, is therefore an empirical question.

This framework guides the analysis below. We first ask whether PACE increases selective-college graduation and improves labor-market outcomes, or instead induces enrollment paths that end without a degree, with weak or negative earnings effects. We then ask whether any negative effects on outcomes are consistent with mismatch. Because preferential admission expands students' choice sets as rejecting a preferential admission is always an option, it can make students worse off ex ante only in the presence of decision-making frictions (Arcidiacono et al., 2011), such as when students misperceive

[15] For a related critique, see Rothstein and Yoon (2008), who revisit Sander's law-school mismatch argument. They argue the evidence points to possible mismatch only among students with the weakest entering academic preparation, but caution that impacts on college performance for this group are hard to interpret because admission preferences also affect their college enrollment.

the probability of completing the degree or returns from the new options. We use students' beliefs about their own graduation chances to measure this friction directly and test the mismatch hypothesis.

5.2 Students in the top 15% of baseline GPA

Higher education outcomes: increased selective-college graduation and continued enrollment in higher education after seven years. Panel A of Table 3 reports our main education outcomes over the seven years after high school graduation: ever enrolling, graduating, remaining enrolled without a degree, and dropping out, separately for selective colleges (columns 1-4) and for any higher education institution (columns 5-8).[16]

PACE did not significantly increase the probability of ever enrolling in any higher education institution during the seven years after high school graduation. However, column 1 shows that PACE increased selective-college enrollment by 12.5 percentage points, a 29 percent increase relative to the control-group mean. The results are consistent with PACE shifting students toward selective colleges. While some of the additional selective-college enrollment resulted in dropout (+3.9 percentage points) and in continued enrollment without a degree by year 7 (+3.5 percentage points), it also translated into a 4.5 percentage point increase in selective-college graduation, or 24 percent relative to the control-group mean. All these results are significant at least at 5 percent level.

For higher education as a whole, PACE did not increase dropout, suggesting that the rise in selective-college dropout reflects, at least in part, a shift in the type of institutions from which students drop out. The main margin of adjustment is continued enrollment: PACE increased the probability of being still enrolled in any higher education institution without a degree by 3.2 percentage points, a 20 percent increase relative to the control-group mean, statistically significant at the 10 percent level. These patterns suggest that PACE moved students into selective colleges and ultimately led more of them to complete selective-college degrees, but also led some students onto higher education trajectories that had not resulted in graduation by year 7.

[16] Appendix B.1 describes how outcome variables are constructed. We define "still enrolled" as having enrolled at least once in the relevant type of institution within seven years after high school graduation, not having obtained a degree by year 7, and being enrolled in year 7. For outcomes measured in any higher education institution, graduation, still enrollment, and dropout are mutually exclusive and collectively exhaustive among students who ever enroll; hence, in the RCT estimates, the treatment effect on ever enrollment equals the sum of the treatment effects on these three outcomes, up to rounding. For selective colleges, this decomposition need not hold because some students enroll in a selective college and later transfer to another type of higher education institution. These students contribute to selective-college ever enrollment but are not classified as selective-college graduates, dropouts, or still enrolled in selective colleges. In the RDD estimates below, the same mechanical decomposition need not hold exactly because the optimal bandwidth can vary across outcomes.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
	A. Main results							
Treatment	0.125***	0.045**	0.035**	0.039***	0.012	-0.029	0.032*	0.008
	(0.029)	(0.022)	(0.017)	(0.014)	(0.014)	(0.024)	(0.019)	(0.016)
Total obs.	2437	2437	2437	2437	2437	2437	2437	2437
Mean	0.426	0.190	0.110	0.074	0.865	0.501	0.164	0.200
	B. Heterogeneity by gender							
Treatment – Women	0.171***	0.090***	0.056***	0.032*	0.032*	0.001	0.033	-0.004
	(0.028)	(0.026)	(0.020)	(0.018)	(0.018)	(0.031)	(0.024)	(0.022)
Mean Women	0.440	0.222	0.107	0.062	0.883	0.556	0.166	0.163
Total obs.	1369	1369	1369	1369	1369	1369	1369	1369
Treatment – Men	0.069	-0.003	0.008	0.046**	-0.011	-0.066*	0.029	0.026
	(0.044)	(0.038)	(0.025)	(0.022)	(0.021)	(0.038)	(0.026)	(0.028)
Mean Men	0.409	0.149	0.115	0.089	0.841	0.432	0.161	0.249
Total obs.	1068	1068	1068	1068	1068	1068	1068	1068
p-value Treat Women=Men	0.020	0.031	0.124	0.627	0.117	0.148	0.894	0.409

Notes: This table reports estimates of coefficient β from regression equation (1). Standard errors clustered at the school level are shown in parentheses. Panel A reports the main results for the full sample. Panel B reports results separately for women and men. At the bottom of the panel, we report the p-value for the test of equality of treatment effects between women and men. *Mean* is the average value of the outcome variable in the control group for the relevant sample. The regressions include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school-region fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

Labor-market outcomes: short-term earning losses and no predicted longer-term gains on average. Panel A of Figure 5 shows the average effects of PACE on formal-sector annual earnings and monthly income for students in the top 15% RCT sample; the corresponding estimates are reported in Table A8. We observe labor-market outcomes through year 7 after high school graduation and predict outcomes for years 8–11 and 12–15 using the surrogate-index procedure described in Appendix B.2. Table A9 reports the corresponding effects on labor supply, measured as months worked and as an indicator for working at least one month in the year.

PACE reduced average annual earnings during the first seven years after high school graduation by 252 thousand pesos, an 11 percent decline relative to the control-group mean. This short-run reduction is consistent with the education results above and with our conceptual framework. PACE shifted students toward selective colleges and increased the probability that they remained enrolled in higher education without a degree by year 7. Because these institutions offer longer programs and are more academically demanding, treated students may have reduced work in the short run in order to devote more time to college, especially while adapting to selective-college coursework. Table A9 supports this

interpretation: treated students worked 0.37 fewer months per year on average in years 1–7, a 10 percent decline relative to the control-group mean.

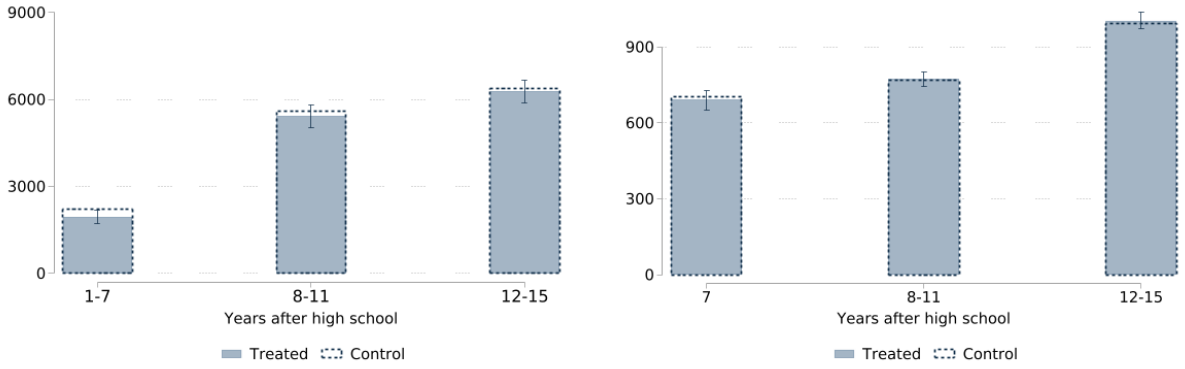
However, PACE had no statistically significant effect on annual earnings in year 7, nor on predicted annual earnings in years 8–11 or 12–15. It also had no statistically significant effect on monthly income conditional on working at any horizon.[17] These patterns suggest that, on average, PACE reduced short run labor supply while some students remained enrolled in more demanding higher education programs, but did not lead to larger future income and earnings on average.[18]

[17]We report average monthly income conditional on working in year 7 to capture any wage impacts in the latest year of our observed data. Table A10 shows sample balance across treatment and control groups, conditional on being employed seven years after high school graduation. The sample is balanced across treatment and control groups in pre-determined characteristics.

[18]All results presented in this section are robust to estimating the specifications without control variables. As shown in Tables A11, A12 and A13, the results are qualitatively and quantitatively similar.

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

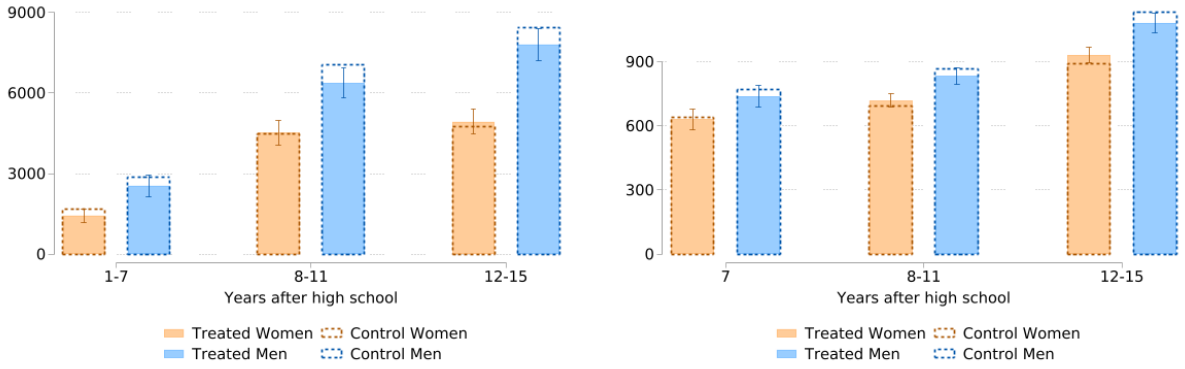
Panel A: All sample



(a) Average yearly earnings

(b) Average monthly income

Panel B: By gender



(c) Average yearly earnings, by gender

(d) Average monthly income, by gender

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, according to GPA in grades 9 and 10. The graphs plot estimates of parameter β in equation (1) and outcome levels. The x-axis indicates the period over which the outcome is averaged: short, medium, or long term. Graphs (a) and (c) use average yearly earnings as the dependent variable, with Graph (c) reporting estimates by gender. Graphs (b) and (d) use average monthly income conditional on working as the dependent variable, with Graph (d) reporting estimates by gender. For average monthly income conditional on working, the short-term impact is reported for year 7 rather than averaged over years 1–7, because year 7 provides a more balanced comparison between treatment and control groups. The bars represent 95% confidence intervals for β calculated from standard errors clustered at school level. 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) and school-region fixed effects. The corresponding estimates are reported in Table A8. The corresponding results for labor force participation are reported in Table A9.

Who benefits? The absence of average medium- and longer-run labor-market effects may mask heterogeneity across students. On average, PACE increased selective-college graduation, a margin that could raise later earnings, but it also increased the probability of being still enrolled in higher education without a degree by year 7, with ambiguous longer-

run implications depending on the (unobserved) counterfactual paths of these students.[19] This raises the question of which students convert expanded access into selective-college degrees, and whether these degree gains are associated with labor-market benefits. We examine this question by studying heterogeneity by gender.

Panel B of Table 3 shows that the gains in selective-college attainment are concentrated among women. PACE increased women's enrollment in selective colleges by 17.1 percentage points and their graduation from selective colleges by 9.0 percentage points, a 41 percent increase relative to the control-group mean statistically significant at 1 percent level. For men, by contrast, the effect on selective-college enrollment is smaller and statistically insignificant, and the effect on selective-college graduation is essentially zero. The gender difference is statistically significant for both selective-college enrollment and selective-college graduation. PACE also increased women's probability of enrolling in higher education overall, although by less than the increase in selective-college enrollment, suggesting both extensive and intensive margin effects. For men, the main statistically significant education effects are an increase in selective-college dropout and a decline in graduation from any higher education institution. Together with the lack of a positive extensive-margin effect on higher education enrollment for men, this pattern suggests that PACE may have redirected some men who would otherwise have completed degrees by year 7, possibly in less selective programs, toward trajectories that had not produced a degree by that time. However, the positive point estimate for the "still enrolled" outcome in column 7, though not statistically significant, suggests that delayed graduation, rather than degree loss, could eventually account for nearly half of men's decline in graduation by year 7.[20]

Table A14 shows that women's graduation gains are not confined to a single field. PACE increased women's graduation from selective colleges in STEM fields by 5.3 percentage points and in non-STEM fields by 3.7 percentage points. For men, PACE increased STEM enrollment in selective colleges but did not significantly increase STEM graduation, while decreasing non-STEM graduation from higher education overall. Thus, among

[19] Absent PACE, they may have never enrolled in higher education, enrolled and dropped out by year 7, or enrolled and graduated by year 7. In the first two cases, since continued enrollment may keep open the possibility of eventual completion, the positive effect on still enrolled could eventually lead to additional degrees, despite potentially delaying labor-market entry; in the third, it may reflect delayed completion, possibly toward a different or more selective degree, or a loss of degree attainment if the student eventually drops out.

[20] At the point estimates, the 6.6 percentage-point decline in men's graduation from any higher education institution by year 7 is mostly offset by similarly sized increases in still enrollment without a degree and dropout; the former may reflect delayed completion and could still lead to a degree, whereas the latter is more likely to represent a loss of degree attainment. Appendix Figure A6 shows how the gender-specific enrollment and graduation trajectories evolve over time in the top 15% RCT sample.

students in the top 15% RCT sample, the additional selective-college degrees generated by PACE are driven by women, in both STEM and non-STEM fields.[21]

Panel B of Figure 5 and Panel B of Table A8 show corresponding gender differences in labor-market outcomes. Among students in control schools, women earn 58.6 cents for every dollar earned by men during years 1–7, close to the national average of 58 cents (World Economic Forum, 2023). This gap is partly explained by labor supply: women work about 30 percent fewer months than men (Table A9).

Three patterns emerge. First, for women, PACE reduced annual earnings during the first seven years after high school graduation, consistent with lower labor supply as more women enrolled in higher education. Table A9 supports this interpretation: women worked fewer months on average over years 1–7, but worked a similar number of months in year 7, when many students begin transitioning into the labor market after graduation. PACE had no effect on women's extensive-margin employment at any horizon. Second, women's longer-run labor-market outcomes are more favorable: PACE increased predicted annual earnings in years 12–15 by 4 percent, although this effect is not statistically significant, and increased monthly income by about 5 percent. This suggests that women's foregone earnings in the short term may be offset over the lifecycle by subsequent income gains. Third, for men, the pattern is clearly less favorable. PACE reduced annual earnings by 15 percent in year 7, by 10 percent in years 8–11, and by 8 percent in years 12–15. It also reduced predicted monthly income in years 12–15 by 4.4 percent. These effects partly reflect lower months worked in year 7 and in the predicted medium and long run, as well as lower long-run employment (Table A9). As a result of these gendered impacts, we predict that the gender gap in monthly income in years 12-15 will be reduced by 38% (from 241.4 to 91.7 thousand pesos).

The labor-market patterns are consistent with the education results: women are the group for whom increased selective-college enrollment translates into additional degrees, while men experience weaker degree gains. These results relate to the literature on gender differences in preferential admission, but show stronger heterogeneity. For example, Black, Denning, and Rothstein (2023) find that the Texas Top Ten Percent policy increased BA completion by 6 percentage points for women induced to attend and by only 2 percentage points for men, with a statistically significant gender difference. In their setting, earnings gains 9–11 years after high school graduation are qualitatively larger for women than for

[21] As selective colleges typically admit students with higher test scores than non-selective institutions, columns 1 and 2 of Table A15 show that PACE also increased the selectivity of the first higher education program attended, measured by the average baseline test scores of peers. Lee bounds imply an increase in selectivity of 0.18–0.23 standard deviations overall, with larger bounds for women than for men. PACE also increased the distance between the average test scores of higher education peers and students' own baseline test scores (columns 3 and 4), and lowered treated students' relative rank within their programs (columns 5 and 6).

men, but the difference disappears in the longer run.[22] Students in our context are more disadvantaged than those affected by the Texas Top Ten Percent policy (Tincani, Kosse, and Miglino, 2026), 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 extends this insight beyond childhood and adolescence: even among disadvantaged students who become top performers by the end of high school, men are less likely than women to improve their long-term mobility when offered additional opportunities to attend selective colleges.

5.3 Marginally eligible students: RDD results

Next, we investigate the effects of preferential admission on lower-performance students around the cutoff for the assignment of PACE slots.

Higher education outcomes: higher selective-college graduation, but higher dropout from higher education. Panel A of Table 4 shows the RDD estimates for students close to the PACE eligibility cutoff, reporting the estimation of the second stage RDD (equation 2).[23] PACE admission offers increased enrollment in selective colleges by 18.2 percentage points and graduation from selective colleges by 7.8 percentage points, a 47 percent increase relative to the mean just below the cutoff. Both estimates are statistically significant at 1 percent level. These results show that, even among marginally eligible students, PACE admission achieved its direct goal of expanding access to and completion in selective colleges.[24]

However, PACE admission also increased dropout from selective colleges by 13.4 percentage points. For higher education as a whole, PACE admission increased the probability of ever enrolling in any institution by 6.7 percentage points, but did not significantly increase graduation from any higher education institution. Instead, it increased dropout from any higher education institution by 9.0 percentage points, a 43 percent increase relative to the mean just below the cutoff. Thus, among marginal students, the increase in selective-college graduation is accompanied by an even larger increase in enrollment followed by dropout from higher education overall: there are 1.2 additional higher education dropouts for each additional selective-college graduate. This pattern suggests that

[22] Antman, Duncan, and Lovenheim (2024) find that affirmative action bans in the United States had negative impacts for targeted women, while for men results were mixed, including positive labor-market outcomes for Black men.

[23] Figure A7 shows the corresponding reduced-form discontinuities in higher education outcomes around the PACE eligibility cutoff.

[24] The education results are robust to estimating the RDD specification without controls. As shown in Table A16, the estimates are qualitatively similar: PACE eligibility increases enrollment and graduation in selective colleges, and increases dropout from selective colleges and from any higher education institution.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
				Panel A: Main results				
Robust	0.182***	0.078***	-0.009	0.134***	0.067***	0.013	-0.037	0.090***
	(0.030)	(0.021)	(0.023)	(0.019)	(0.020)	(0.031)	(0.027)	(0.025)
Bandwidth	35	58	38	58	58	54	37	56
Bandwidth obs.	7257	11518	7817	11518	11518	10834	7571	11233
Mean	0.409	0.167	0.116	0.077	0.856	0.477	0.179	0.207
				B. Heterogeneity by gender				
Robust - Women	0.167***	0.113***	-0.051*	0.136***	0.064***	0.053	-0.080**	0.091***
	(0.041)	(0.027)	(0.027)	(0.025)	(0.023)	(0.039)	(0.032)	(0.032)
Bandwidth	34	59	39	46	51	57	32	49
Bandwidth obs.	4146	7039	4793	5565	6039	6717	3919	5815
Mean	0.428	0.165	0.117	0.065	0.873	0.517	0.181	0.186
Robust - Men	0.222***	0.029	0.052*	0.142***	0.028	-0.087**	-0.002	0.076**
	(0.035)	(0.034)	(0.030)	(0.032)	(0.031)	(0.043)	(0.038)	(0.038)
Bandwidth	68	49	58	55	72	68	47	71
Bandwidth obs.	5486	3980	4687	4486	5799	5486	3864	5776
Mean	0.406	0.154	0.112	0.097	0.833	0.415	0.179	0.236
p-value Women=Men	0.185	0.181	0.007	0.854	0.912	0.137	0.114	0.783
				C. Heterogeneity by overconfidence based on graduation beliefs				
Robust - High	0.199	-0.028	0.001	0.296***	0.035	-0.026	0.019	0.324**
	(0.176)	(0.155)	(0.105)	(0.111)	(0.110)	(0.170)	(0.135)	(0.145)
Bandwidth	61	47	67	49	64	44	62	41
Bandwidth obs.	350	269	374	282	357	255	350	241
Mean	0.505	0.217	0.145	0.079	0.926	0.487	0.188	0.156
Robust - Low	0.291	0.130	0.011	0.214	0.001	0.227	0.017	-0.168
	(0.238)	(0.172)	(0.169)	(0.233)	(0.143)	(0.281)	(0.152)	(0.264)
Bandwidth	69	51	51	43	93	53	58	49
Bandwidth obs.	266	199	201	171	353	211	227	192
Mean	0.287	0.084	0.100	0.150	0.768	0.301	0.171	0.274
p-value High=Low	0.717	0.559	0.929	0.700	0.873	0.521	0.857	0.086

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. Panel A reports the main results for the full sample. Panel B reports results separately for women and men. At the bottom of the panel, we report the p-value for the test of equality of treatment effects between women and men. Panel C reports results separately for students with above- and below-median overconfidence about their graduation probability. At the bottom of the panel, we report the p-value for the test of equality of treatment effects between these two groups. *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 standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, and high school track (academic or vocational). Panel A and Panel B additionally include school fixed effects; Panel C excludes school fixed effects.

preferential admission generated additional selective-college degrees for some marginal students, but also induced other students into higher education trajectories that led to dropout within 7 years.[25]

[25] As in the RCT analysis, the outcomes for any higher education institution are collectively exhaustive and mutually exclusive among students who ever enroll. In the RDD estimates, however, the treatment effects across columns need not add up exactly because the optimal bandwidth can differ across outcomes.

The mixed education effects occur alongside evidence that PACE moved marginal students into more selective, and potentially more demanding, programs. Because program selectivity and within-program rank are observed only for students who enroll, Table A18 reports Lee bounds for these outcomes. The estimates show that PACE increased the selectivity of the first higher education program attended by 0.13–0.26 standard deviations, increased the distance between students' own baseline test scores and the average baseline test scores of their higher education peers, and lowered students' relative rank within their higher education programs. These patterns are consistent with both sides of the conceptual framework: marginal students gained access to higher-quality programs, but they also faced more academically demanding environments.

Labor market outcomes: no average impacts on earnings or monthly income. Panel A of Figure 6 shows the RDD estimates for formal-sector annual earnings and monthly income conditional on working; the estimates are reported in Table A17.[26] Table A19 reports the effects on labor supply, measured by months worked and by an indicator for working at least one month in the year. PACE admission offers have no statistically significant average effect on annual earnings in year 7, over years 1–7, or in the predicted medium- and long-run horizons. They also have no statistically significant average effect on monthly income conditional on working at any horizon. Similarly, Table A19 shows no statistically significant effects on months worked in year 7 or in the predicted medium- and long-run horizons, and only a small decline in average employment over years 1–7. [27]

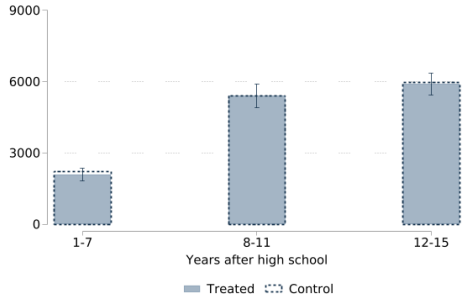
Who benefits? The absence of effects on earnings or monthly income on average is consistent with the mixed education results. Preferential admission increased selective-college graduation for some marginal students, a margin that could raise later earnings, but also increased dropout from higher education for others, a margin that could reduce labor supply while enrolled without generating degree returns. These opposing channels may offset each other on average. The key question, therefore, is not whether marginal students benefit on average, but which students obtain the additional selective-college degrees and which students experience the increase in dropout. We examine this question focusing first on heterogeneity by gender.

[26]Figure A8 shows the corresponding reduced-form (ITT) discontinuities for observed labor-market outcomes in year 7 and over years 1–7. Furthermore, Table A20 reports balance tests for predetermined characteristics in the subsample observed working seven years after high school graduation. The estimated discontinuities are small and statistically insignificant for all covariates except one, indicating that predetermined characteristics remain continuous at the PACE eligibility cutoff even after conditioning on employment.

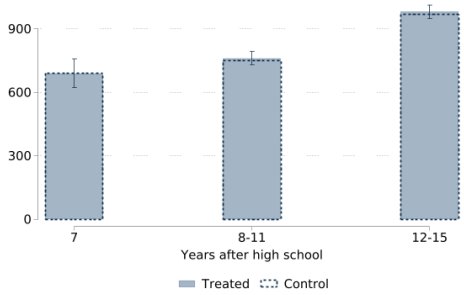
[27]The labor-market results are robust to estimating the RDD specification without controls while retaining school fixed effects. As shown in Tables A21 and A22, the estimates are qualitatively similar: PACE admission offers do not generate statistically significant average effects on earnings or monthly income, and the effects on months worked and employment remain small and mostly insignificant.

Figure 6: Labor market effects – RDD sample

Panel A: All sample

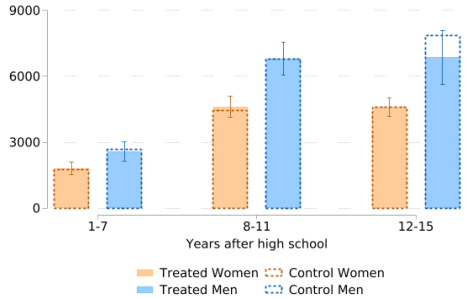


(a) Average yearly earnings

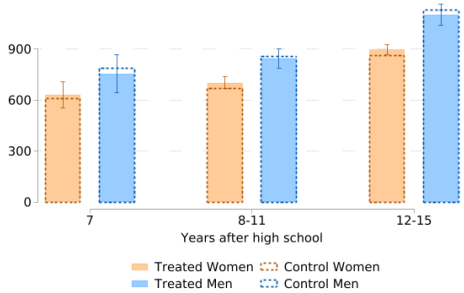


(b) Average monthly income

Panel B: By gender

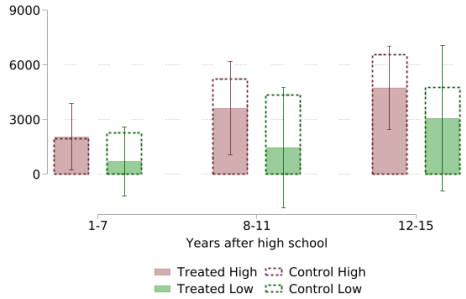


(c) Average yearly earnings, by gender

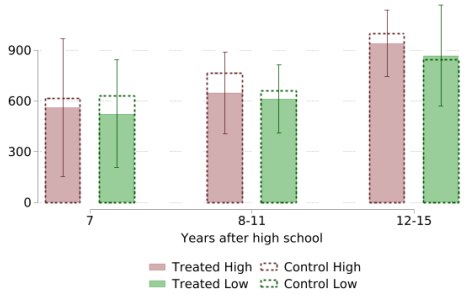


(d) Average monthly income, by gender

Panel C: By overconfidence



(e) Average yearly earnings, by overconfidence



(f) Average monthly income, by overconfidence

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 graphs plot estimates of parameter δ in equation (2) and outcome levels. The x-axis indicates the period over which the outcome is averaged: short, medium, or long term. Graphs (a), (c) and (e) use average yearly earnings as the dependent variable, with Graph (c) reporting estimates by gender and Graph (e) by overconfidence in graduation probability. Graphs (b), (d) and (f) use average monthly income conditional on working as the dependent variable, with Graph (d) reporting estimates by gender and Graph (f) by overconfidence in graduation probability. For average monthly income conditional on working, the short term impact is reported for year 7 rather than averaged over years 1–7, because year 7 provides a more balanced comparison across students around the cutoff. The bars represent 95% confidence intervals for δ calculated from standard errors clustered at school level. The regressions include the standard controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade and high school track (academic or vocational). Panel A and Panel B additionally include school fixed effects; Panel C excludes school fixed effects. The corresponding estimates are reported in Appendix Table A17. The corresponding results for labor force participation are reported in Appendix Table A19.

Panel B of Table 4 shows that PACE admission offers increased selective-college enrollment for both women and men. The post-enrollment trajectories differ, however. For women, PACE increased graduation from selective colleges by 11.3 percentage points. PACE also increased women's enrollment in any higher education institution and dropout from both selective colleges and any institution, while reducing the probability of remaining enrolled in higher education without a degree by year 7, both in selective colleges and in any institution. This pattern suggests that, for women, preferential admission not only generated additional selective-college degrees, but may also have shortened unresolved education trajectories, either by accelerating completion or by reducing the time spent enrolled before dropout (Appendix Figure A9). For men, the effect on selective-college graduation is smaller at the point estimate and not statistically significant. Men instead experienced a statistically significant decline in graduation from any higher education institution, by 8.7 percentage points, together with an increase in dropout from higher education. In contrast to the RCT results, where part of men's decline in graduation by year 7 may reflect delayed completion, the RD estimates show no offsetting increase in still enrollment at any institution: the point estimate in column 7 is essentially zero. Thus, the reduction in men's degree attainment by year 7 in the RD appears less consistent with delayed completion and more consistent with a loss of degrees. These patterns suggest that marginal women are more likely to convert the additional selective-college access into selective-college degrees and non-completing trajectories appear to resolve more quickly, while marginal men are more likely to experience degree losses.

Table A23 shows that the selective-college graduation gains are concentrated in non-STEM fields. In the full RDD sample, PACE increased graduation from selective colleges in non-STEM fields by 8.5 percentage points, while it did not increase selective-college graduation in STEM fields. By gender, both women and men experienced gains in non-STEM selective-college graduation. For men, however, these gains were accompanied by lower selective-college graduation in STEM fields, helping explain why the overall effect on men's selective-college graduation is small and statistically insignificant. PACE also reduced men's STEM graduation from any higher education institution and increased their dropout from STEM programs in selective colleges, while increasing their enrollment in non-STEM programs. These patterns are consistent with PACE shifting some marginal men away from STEM completion and toward non-STEM enrollment paths, some of which had not produced a degree by year 7. Taken together, these patterns suggest that the additional selective-college degrees generated at the margin come primarily from non-STEM programs, while some of the adverse completion effects for men appear in STEM fields.

Panel B of Figure 6 and of Tables A17 and A19 show the corresponding labor-market patterns. For women, PACE admission offers do not significantly affect annual earnings,

but they increase monthly income in the predicted medium- and long-run horizons by a small but statistically significant impact of 4-5%. The effects on women's labor supply are small, with a decline in average employment over years 1–7 while in education but no persistent reduction in months worked or employment.[28] For men, the labor-market pattern is less favorable: PACE does not increase monthly income conditional on working, with negative but insignificant impacts at all horizons, and significantly reduces predicted annual earnings in years 12–15 by 12%. These results are consistent with the education patterns. Women are the group for whom selective-college access more clearly translates into additional selective-college degrees and higher monthly income among workers, while men experience non-STEM selective-college degree gains that are accompanied by STEM degree losses, losses in graduation from any higher education institution, and less favorable labor-market outcomes.

Information frictions. Why would some marginal students accept preferential admission offers to programs from which they are likely to drop out? One possibility is that students gain enough utility from time spent in selective colleges to offset any costs associated with dropping out. In this case, enrolling and dropping out may still be welfare-improving. An alternative possibility is that students do not fully anticipate the academic demands of selective colleges, so that dropout may reflect errors in enrollment decisions. This mechanism is central to the mismatch hypothesis (Arcidiacono et al., 2011), but is difficult to test directly without data on students' beliefs. We use survey data collected from students in the RCT schools during their final year of high school, linked to administrative records on their subsequent higher education outcomes.[29]

Figure A10 shows that students are highly optimistic about their chances of completing a selective-college degree if admitted.[30] The modal response is that students are certain they would graduate from a selective college, and this response is even more common among students who eventually enroll in one. Contrasting high school students' perceived and actual chances of graduating from a selective college reveals a marked overoptimism regarding future college performance. Among surveyed students who enrolled in a selective college, only 39.7 percent graduated within 7 years (Table A25). Even among students who reported being certain that they would graduate, only 41.8 percent did so. Figure 7 compares perceived graduation probabilities with predicted actual graduation probabilities and shows that students are overoptimistic throughout the distribution of

[28]Despite the negative impact on labor force participation in years 1-7, the effect for women's earnings is almost zero and insignificant in the same period suggesting some potential positive effects on their wages.

[29]Table A24 compares students with and without non-missing overconfidence measures in the RDD sample. The two groups are broadly balanced on predetermined characteristics.

[30]Over-optimism about graduation chances was documented also among Berea college students in the United States (Stinebrickner and Stinebrickner, 2014).

actual completion chances.[31] Table A26 further shows that overconfidence is substantially higher among men: women are about 0.3 standard deviations less overconfident than men, both with and without controls. The gender gap in overconfidence may be a key driver of differential effects 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 provides suggestive evidence that they are. Among students with above-median overconfidence, PACE admission offers increase dropout from any higher education institution by 32.4 percentage points. Among students with below-median overconfidence, the corresponding estimate is negative and statistically insignificant. The difference between the two estimates is statistically significant at the 10 percent level.[32] Thus, the increase in dropout generated by preferential admission is driven primarily by students who were more overconfident about their probability of completing selective college.

Table A27 helps interpret this pattern. Among high-overconfidence students, PACE substantially increases the distance between students' own baseline test scores and the average baseline test scores of their peers in the higher education programs they attend. The corresponding bounds are smaller and statistically insignificant among low-overconfidence students. This suggests that overconfident students were more likely to use PACE to sort into programs for which they were relatively less academically prepared, consistent with the mechanism posited by the mismatch hypothesis. Panel C of Figure 6 and the corresponding appendix tables show noisier labor-market patterns by overconfidence: low-overconfidence students work fewer months on average over years 1–7, while high-overconfidence students are predicted to work fewer months in years 12–15. However, this evidence is based on a limited number of observations, since the survey data were only collected for one cohort and only in schools that were part of the randomized expansion.[33] Nonetheless, taken together, the education results support the view that preferential admission generated mismatch at the margin, with information frictions contributing to it: the dropout increase is concentrated among students who overestimated

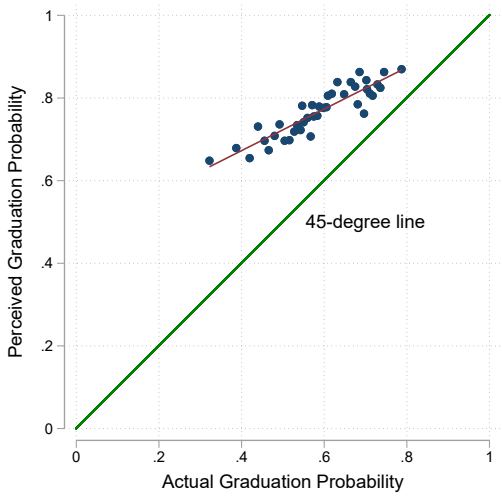
[31] Figure A11 shows that this pattern is robust to an alternative numerical coding of the survey responses.

[32] The same pattern appears for selective-college dropout: the point estimate is positive and statistically significant among high-overconfidence students, but smaller and statistically insignificant among low-overconfidence students.

[33] 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. In contrast, the beliefs are baseline measures in the RDD identification strategy as they were collected a few months before college admissions.

their chances of completing a selective-college degree, consistent with enrollment decisions based on mistaken expectations.

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 probability of graduating within seven years (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 A11 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 seven 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.

6 Linking the RCT and RDD results

The RCT and RDD analyses answer complementary questions, but they differ along two dimensions. The RCT estimates the effect of attending a PACE school for students targeted by the policy, combining the possibility of receiving a preferential admission offer with other channels of exposure to PACE during high school. The RDD instead isolates the effect of preferential admission for students in PACE schools who are marginal for eligibility. The contrast between the two sets of results is informative. In the top 15% RCT sample, PACE increases selective-college graduation without increasing dropout from higher education overall. At the admission margin, preferential admission also increases selective-college graduation, but this gain is accompanied by a larger increase in dropout from any higher education institution, with evidence suggesting that the dropout effect is consistent with mismatch. This contrast raises two questions. First, at what point in the within-school academic ranking does the access-mismatch tradeoff emerge? Second,

how much of the RCT effect of attending a PACE school reflects the preferential-admission channel, rather than other channels through which PACE schools may affect students?

In this section, we use the coexistence of the RCT and RDD in the same policy setting to answer these questions in two steps. We first extrapolate away from the admission cutoff to study how the effect of preferential-admission eligibility varies along the within-school academic ranking (Section 6.1). We then use a potential-outcomes framework to decompose the RCT estimand into the component attributable to preferential-admission eligibility for eligible students and the residual component attributable to other PACE-school channels (Section 6.2). The result relies on the availability of both experimental and discontinuity-induced variation in the same empirical setting. It provides a transparent interpretation of how the two estimates relate to each other and clarifies the condition under which the RCT estimate can serve as an experimental benchmark for assessing the validity of the RDD extrapolation method, in the spirit of LaLonde (1986).

6.1 Extrapolating RDD results away from the cutoff

The RDD estimates in Section 5.3 identify the effects of preferential admission for students close to the PACE eligibility cutoff, that is, for the lowest-ranked students among those eligible for the admission offer. These estimates are central for evaluating the marginal student admitted under the current rule, but they do not directly answer a broader policy question: how far down the within-school academic ranking can preferential-admission eligibility be extended before a trade-off emerges between selective-college degree gains and higher dropout? To address this question, we follow Angrist and Rokkanen (2015) and extrapolate away from the cutoff.[34] In this exercise, we focus on the effect of preferential-admission eligibility rather than preferential admission itself, which is offered only to eligible students who apply. Eligibility is more policy-relevant as it is the margin directly set by policy.

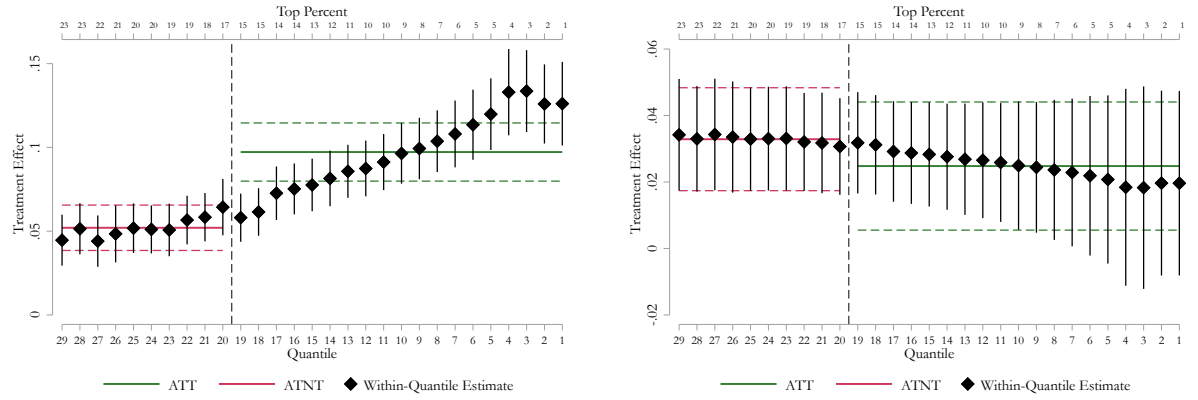
Following Angrist and Rokkanen (2015), we use pre-treatment (i.e., pre-admission) covariates to predict potential outcomes away from the cutoff. The key identifying assumption is that, conditional on these covariates and high school fixed effects, potential outcomes are mean independent of the running variable. We also require common support, so that eligible and ineligible students with comparable covariates are observed over the relevant range of the running variable. Under these assumptions, the estimated treatment effect for students away from the cutoff is obtained by comparing conditional mean

[34]This approach has been used in related settings to estimate treatment effects for inframarginal units away from an RD cutoff; see, for example, Cingano et al. (2025). Because PACE eligibility is determined by school-specific cutoffs, as in the RDD analysis in the previous Section we normalize each student's running variable by subtracting the relevant school-specific cutoff and we include in our specification high school fixed effects, following the logic in Fort et al. (2022) for multi-cutoff RD designs.

outcomes for eligible and ineligible students with the same covariates and averaging these differences over the covariate distribution in each target interval of the running variable. Appendix B.4 empirically assesses the validity of these assumptions, showing that they are reasonable in our setting, and presents robustness of the results to varying the set of covariates.

Figure 8 reports the extrapolated effects for the two outcomes that summarize the main access–mismatch tradeoff: graduation from a selective college and dropout from any higher education institution. Panel A shows that the effect on selective-college graduation remains positive as we move away from the cutoff, including among students who rank below the current top 15% cutoff. Panel B, however, shows that positive dropout effects already emerge among students ranked above the current top 15% admission margin. Within the population of students who take the entrance exam, the first statistically detectable increase in dropout from any institution appears once eligibility is extended past roughly the top 6 percent of the within-school GPA distribution.[35]

Figure 8: RDD away from the cutoff



(a) Graduation from a selective college

(b) Dropout from any institution

Notes: This figure reports estimates of the effects of preferential-admission eligibility away from the cutoff, following Angrist and Rokkanen (2015). The conditioning vector consists of grade-12 GPA. ATT refers to the average treatment effect on eligible students, and ATNT to the average treatment effect on ineligible students. Vertical bars correspond to 95% confidence intervals, obtained from standard errors clustered at school level computed using a nonparametric bootstrap with 100 replications. The lower horizontal axis indexes quantiles of the running variable distribution. The upper horizontal axis reports, for each group, the average school-wide GPA ranking, calculated among all students in the school and based on GPA in all four high school years. These rankings are rounded to the nearest integer, so two distinct groups may display the same integer cutoff when their exact values fall in the same rounding interval.

Our extrapolation suggests that, up to roughly the top 6% within-school GPA cutoff, preferential-admission eligibility generates additional selective-college degrees for some

[35]The PACE running variable is available to us only for students who take the entrance exam. We therefore conduct the extrapolation in this sample and map the estimated threshold back into the within-school GPA distribution; Appendix B.4 describes this conversion. Interpreting this mapped threshold as a policy counterfactual for the full school cohort also requires assuming that changing the eligibility cutoff would not change entrance-exam taking as a function of within-school GPA rank, and that non-exam takers at a given rank would have persistence outcomes similar to exam takers at that rank.

students without a statistically detectable increase in enrollment followed by dropout for others. Beyond this cutoff, however, a tradeoff between these outcomes emerges. This cutoff corresponds roughly to the 60^{th} percentile of the national distribution of grade-10 test scores; although the cutoff targets the top 6% within schools, PACE schools serve relatively low-achieving students, therefore this cutoff reaches much farther down the national distribution.[36] The results suggest that, even without diluting course content or lowering grading standards, preferential admissions could likely be extended to roughly the top 40% of students in terms of high school test scores without generating mismatch-related dropout. However, this threshold should be interpreted as an approximate benchmark rather than an exact rule as we discuss in Appendix B.4.

6.2 Decomposing RCT results

In this section we use the potential-outcome framework to show how estimates from the RCT and RDD-extrapolation analyses can be functionally linked, decomposing the RCT impacts into a component due to preferential-admission eligibility and one due to other channels. To satisfy key assumptions in the derivations, the empirical results focus on the experimental sample only, and consider all students, regardless of baseline within-school GPA. Accordingly, this exercise identifies the contribution of non-admission PACE-school channels for the full experimental sample, rather than directly decomposing the top-15% RCT estimate; it is informative about the latter only if these channels have the same average effect among top-15% students as in the full sample.

6.2.1 Theoretical results

Potential outcomes and treatments The RCT analysis estimates the average treatment effect of being in a PACE school. The RDD-extrapolation analysis, implementing the method developed by Angrist and Rokkanen (2015), provides an estimate of the average effect of preferential admission eligibility for all those eligible—i.e., above the top 15% cutoff in their PACE school and who took the PSU.[37] There are, therefore, two treatments:

– being in a PACE school

[36] Appendix Figure B.3 reports extrapolated effects of eligibility on monthly income in year 7 and annual earnings in years 1–7 after high school. These effects are imprecisely estimated and almost never statistically distinguishable from zero.

[37] More precisely, the method estimates the effect of eligibility defined as taking the PSU, having a score above the admission cutoff, and having attended the PACE school for two continuous years. That is, the definition of eligibility used to define the treatment is slightly different from the definition used to define the treated population, which ignores the two-years criterion. As nearly all students satisfy this criterion, for notational simplicity we refer to the eligible as those who satisfy the PSU and score requirements.

– being eligible for a preferential admission, i.e., the difference between the running variable and the within-school cutoff is non-missing and positive. The running variable is non-missing if and only if the student takes the PSU (even though the PSU score is not one of its components).[38]

We consider three potential outcomes Y_{sa} that vary depending on whether the student receives the PACE school treatment ($s = 0, 1$) and the eligibility treatment ($a = 0, 1$):

– Y_{11}: PACE school and eligibility

– Y_{10}: PACE school and no eligibility

– Y_{00}: no PACE school and no eligibility

Notice that students cannot get eligibility without the PACE treatment: Y_{01} does not exist.

The potential outcome when receiving the PACE school treatment is

$$Y_{1*} = aY_{11} + (1 - a)Y_{10}. \tag{4}$$

We use the lower case letters s and a to refer to the (theoretical) treatments, and the upper case letters S and A to indicate whether a student in the data is observed receiving each treatment. Therefore, $S = 1$ if a student is observed in a PACE school and $S = 0$ otherwise, and $A = 1$ if a student is observed receiving the eligibility treatment and $A = 0$ otherwise. While no student observed in a non-PACE school ($S = 0$) is ever observed receiving preferential admission eligibility ($A = 1$), it is possible to define the potential outcome under admission eligibility for these students (Y_{11}), as well as the potential likelihood of eligibility had their school been treated ($Pr(a = 1)$).

As in Heckman and Navarro-Lozano (2004), we write the potential outcomes in terms of variables observed by the researcher (X) and unobserved variables specific to each potential outcome, assuming additive separability for simplicity:

$$Y_{11} = \mu_{11}(X) + \epsilon_{11} \tag{5}$$
$$Y_{10} = \mu_{10}(X) + \epsilon_{10} \tag{6}$$
$$Y_{00} = \mu_{00}(X) + \epsilon_{00} \tag{7}$$

We define the following individual-level treatment effects:

[38] While the within-school rank is calculated among all students in the school, the Ministry provides the running variable to researchers only for the subset who took the entrance exam.

– Effect of being in PACE school: $\Delta_s = Y_{1*} - Y_{00}$.

– Effect of eligibility for preferential admission if in PACE school: $\Delta_a = Y_{11} - Y_{10}$.

Lemma 1. *The joint effect of being in a PACE school and being eligible for a preferential admission is the sum of the effect of being eligible for a preferential admission conditional on being in a PACE school and the effect of being in a PACE school without eligibility:*

$$\underbrace{Y_{11} - Y_{00}}_{\substack{\textit{effect of PACE school} \\ \textit{with eligibility}}} = \underbrace{(Y_{11} - Y_{10})}_{\substack{\textit{effect of eligibility} \\ \textit{cond on PACE school}}} + \underbrace{(Y_{10} - Y_{00})}_{\substack{\textit{effect of PACE school} \\ \textit{without eligibility}}} \tag{8}$$

Proof: See Appendix C.

Parameters of interest and decomposition result Define the two parameters of interest, conditional on covariates.

Definition 1. *Conditional average effect of being in a PACE school (CATE$_S$):* $E[Y_{1*} - Y_{00}|X]$.

Definition 2. *Conditional average effect of preferential admission eligibility on the eligible in PACE schools (CATT$_A$):* $E[Y_{11} - Y_{10}|X, S = 1, A = 1]$.

We now present our main decomposition result. We begin by deriving a decomposition of treatment effects conditional on covariates, reported in Theorem 1. We then derive the main result, a decomposition of average, unconditional treatment effects, in Theorem 2.

Assumption 1. *Conditional balance of potential eligibility. The potential likelihood of admission eligibility is the same across PACE and non-PACE schools conditional on X:*

$$Pr(a = 1 \mid X) = Pr(a = 1 \mid X, S), \quad S = 0, 1.$$

Assumption 2. *Conditional mean balance of eligibility effects among potential eligibles. For all X such that $Pr(a = 1 \mid X) > 0$,*

$$E[Y_{11} - Y_{10} \mid X, a = 1, S = 1] = E[Y_{11} - Y_{10} \mid X, a = 1].$$

Assumption 1 is stronger than a standard experimental balance condition because eligibility may also depend on post-randomization behavior. For example, treated students could adjust their GPA in grades 11 and 12 in a way that would affect their probability of crossing the eligibility threshold. The evidence, however, suggests that the main rank component of eligibility is largely predetermined. Table A28 shows lack of treatment impacts on rank in GPA considering all grades from 9 to 12. Consistent with this, Figure

3 shows that only 2.7% of students outside the baseline top 15% of the within-school GPA distribution receive a preferential admission, indicating limited movements across the relevant rank threshold between the experimental baseline and the end of high school. The evidence therefore supports the the plausibility of Assumption 1.

Assumption 2 requires that, among students who would be eligible for preferential admission, the mean effect of eligibility is balanced across PACE and non-PACE schools conditional on X. This assumption is not directly testable because individual eligibility effects are unobserved. However, random assignment and balance in predetermined covariates among students in the top 15% at baseline (Table A6), and the persistence of GPA rank between baseline and the end of high school, support the plausibility of comparing potential eligibles in treated and control schools conditional on X.

Theorem 1. Conditional decomposition. *Under Assumptions 1 and 2, the conditional average treatment effect of being in a PACE school ($CATE_S$) can be decomposed into the sum of two terms: the conditional average effect of eligibility on the eligible ($CATT_A$) multiplied by the conditional probability of eligibility, and the conditional average effect of the PACE school treatment without eligibility:*

$$\underbrace{E[Y_{1*} - Y_{00}|X]}_{CATE_S} = \underbrace{\pi_X}_{\text{Cond. fraction eligible}} \underbrace{E[Y_{11} - Y_{10}|X, A=1, S=1]}_{CATT_A} + \underbrace{E[Y_{10} - Y_{00}|X]}_{\text{Conditional average effect of PACE school without eligibility}} \tag{9}$$

Proof: See Appendix C.

We now proceed to decompose the average effect of being in a PACE school unconditional on covariates. Define the following treatment effect parameters of interest:

Definition 3. *Average treatment effect of being in a PACE school (ATE_S):* $\int E[Y_{1*} - Y_{00}|X]f(X)dX$.

Definition 4. *Average effect of eligibility on the eligible in PACE schools (ATT_A):* $\int E[Y_{11} - Y_{10}|X, A=1, S=1]f(X|A=1, S=1)dX$.

We make the additional assumption of covariate balance in the RCT, for which Table A29 provides supportive evidence:

Assumption 3. *Covariate balance. The distribution of X is equal to the distribution of X conditional on $S=1$ for all values of X:* $f(X) = f(X|S=1)\ \forall X \in \mathcal{X}$.

Theorem 2. Unconditional decomposition. *Under Assumptions 1, 2, and 3, the average treatment effect of being in a PACE school (ATE_S) can be decomposed into the*

sum of two terms: the average effect of eligibility on the eligible (ATT_A) multiplied by the fraction of eligible in PACE schools, and the average effect of the PACE school treatment without eligibility:

$$\underbrace{\int E[Y_{1*} - Y_{00}|X]f(X)dX}_{ATE_S} = \underbrace{\pi}_{Fraction\ eligible}\underbrace{\int E[Y_{11} - Y_{10}|X, A=1, S=1]f(X|A=1, S=1)dX}_{ATT_A} + \tag{10}$$

$$+ \underbrace{\int E[Y_{10} - Y_{00}|X]f(X)dX}_{ATE\ of\ PACE\ school\ without\ eligibility}$$

Proof: See Appendix C.

Under a balance assumption reasonably satisfied in our experimental sample (Assumption 3) and two additional conditional independence assumptions that are plausible in our setting (Assumptions 1 and 2), Theorem 2 shows that the average treatment effect of attending a PACE school, identified in the RCT analysis, can be decomposed as the sum of two terms. The first term, $\pi \cdot ATT_A$, is the product of the share of students who are eligible for a preferential admission in PACE schools and the average effect of eligibility for those students, identified from the RDD-extrapolation analysis. The second term is the average effect of attending a PACE school without receiving PACE admission eligibility, capturing all channels through which PACE schools may affect outcomes other than via the preferential admission itself. This term is not directly estimable from the data, but it can be determined residually from parameters that are: $ATE_S - \pi \cdot ATT_A$.

If we have reasons to believe that PACE schools have no effect on outcomes in the absence of preferential admission eligibility, that is, the PACE-without-admission-eligibility term is zero, then the decomposition in equation (10) implies that the experimental estimate of ATE_S should equal $\pi \cdot ATT_A$ when Assumptions 1, 2, and 3 are satisfied. Under this additional assumption, the RCT estimate therefore provides an experimental benchmark to validate the extrapolated RDD estimate of ATT_A, in the spirit of LaLonde (1986): any discrepancy between ATE_S and $\pi \cdot ATT_A$ in this case would reflect bias in the RDD-extrapolation method of Angrist and Rokkanen (2015).

6.2.2 Empirical results

Table 5 reports the estimates of the parameters in equation (10) for enrollment followed by graduation from a selective college and enrollment followed by dropout from any higher education institution. For each outcome, the table reports the experimental estimate of

ATE_S, the extrapolated estimate of ATT_A, and the implied residual effect of attending a PACE school without obtaining preferential-admission eligibility, $ATE_S - \pi \cdot ATT_A$.[39]

Table 5: Decomposition result, estimates

Outcome	(1)	(2)	(3)
	From RCT: ATE_S	From RDD extrapolation: ATT_A	Residual term: $ATE_S - \pi \cdot ATT_A$
Graduation, selective college	0.015	0.162	-0.009
Dropout, any institution	0.002	-0.003	0.002

Notes: This Table reports estimates of the parameters in equation (10). Column (1) reports the estimates for the average treatment effect of being in a PACE school, ATE_S, obtained as coefficient β in regression equation (1), estimated without controls and in the entire experimental sample (including students of all initial rank levels). Column (2) reports the estimate, obtained by applying the method of Angrist and Rokkanen (2015) to all PACE schools in the experimental sample, of the average treatment effect of eligibility on eligible students whose covariates lie in the common-support region, ATT_A^{CS}. Appendix B.5 explains the condition under which we can interpret ATT_A^{CS} as ATT_A, i.e., the average treatment effect of eligibility on all eligible students. Column (3) reports the residual term, obtained as $ATE_S - \pi \cdot ATT_A$, which represents the impact of attending a PACE school without preferential admission eligibility. Parameter π is estimated from the fraction of eligible students in PACE schools in the sample, which equals 0.148.

Across both outcomes, the residual estimates are close to zero. The evidence therefore suggests that the experimental effects of attending a PACE school are largely accounted for by the preferential-admission eligibility component, $\pi \cdot ATT_A$, with little evidence of an additional effect of PACE-school exposure alone. If we have prior reason to believe that the school-only treatment has zero impacts, then these empirical results are consistent with the validity of the Angrist and Rokkanen (2015) extrapolation method in this setting.

7 Conclusions

A growing literature on affirmative action in higher education studies the tradeoff between gains for targeted students and potential losses for untargeted students (Bertrand, Hanna, and Mullainathan, 2010; Mello, 2022; Black, Denning, and Rothstein, 2023; Otero, Barahona, and Dobbin, 2023; Machado, Reyes, and Riehl, 2025). This paper focuses instead on a tradeoff that can arise within the targeted group itself. Whether preferential admission improves the long-run outcomes of targeted disadvantaged students depends on whether they can convert access to selective colleges into degrees and later labor-market gains. We study this question in the context of PACE (*Programa de Acompañamiento*

[39]For these outcomes, we assessed the conditional-independence assumption and common support underlying the RDD extrapolation, and found them to be reasonable in this analysis' sample. Consistent with the theoretical result, the vectors include only covariates that are measured before both the RCT assignment and the RDD eligibility determination: for dropout, average GPA in grades 9 and 10 interacted with GPA in grade 9, GPA in grade 9 interacted with gender, and age; for selective-college graduation, average GPA in grades 9 and 10 interacted with GPA in grade 9, within school rank based on average GPA in grades 9 and 10 interacted with GPA in grade 10, dummy for having never failed a year interacted with gender, grade-10 SIMCE test score squared, and grade 10-SIMCE test score interacted with age.

y Acceso Efectivo a la Educación Superior), a Chilean top-percent plan that targets the country's most disadvantaged high schools and offers preferential admission to selective colleges to students graduating near the top of their high school cohort. Because PACE reaches students who, while top-performing within their schools, score far below regular college entrants on national high school tests, it provides an unusually demanding setting in which to test how far preferential admission can be extended while still benefiting the students it targets.

Our results show that preferential admission can increase selective-college graduation, but that its benefits depend on which students are reached. Among students in the top 15% of baseline high school GPA, attending a PACE school increased graduation from selective colleges without increasing dropout from higher education overall. Although PACE reduced short-run formal-sector earnings and labor supply, consistent with students spending more time enrolled in demanding colleges, it did not generate statistically significant average labor-market gains at the horizons we predict (8-15 years after high school graduation). But these average effects mask sharp heterogeneity. For women, PACE increased selective-college graduation in both STEM and non-STEM fields and raised predicted longer-run monthly income. For men, by contrast, we do not detect selective-college degree gains; they were less likely to hold a higher education degree by year 7, likely reflecting delayed completion and possible degree losses, and they obtained more unfavorable predicted labor-market outcomes.

For students marginally eligible for the PACE admission offer, a starker tradeoff emerges. The RDD estimates show that preferential admission increased enrollment in and graduation from selective colleges. However, these gains were accompanied by an even larger increase in dropout from higher education, while average effects on formal-sector earnings and monthly income were not statistically significant overall. This pattern suggests that preferential admission generated selective-college degrees for some marginal students while inducing others into higher education trajectories that ended in dropout within seven years. The heterogeneity results again point to women as the group more likely to convert preferential admission into selective-college degrees. Men, by contrast, were less likely to hold a higher education degree by year 7, likely reflecting degree losses rather than delayed completion, and had more unfavorable predicted labor-market outcomes. The results also point to information frictions as an important mechanism behind mismatch at the margin. Students substantially overestimated their probability of graduating from a selective college, and overconfidence was higher among men. The increase in dropout among marginal students was concentrated among those with above-median overconfidence, who were also more likely to enter programs where their own baseline academic preparation was far below that of their peers. Even if based on a small sample, the evidence is consistent with a key tenet of the 'mismatch hypothesis' (Arcidiacono

et al., 2011): decision-making frictions are necessary for affirmative action to make targeted students ex ante worse off. Although prior work has stressed the role of students' beliefs in mismatch, administrative data linked to students' beliefs within an affirmative action policy context are rare (Arcidiacono, Aucejo, and Hotz, 2016).

Extrapolating the RDD estimates away from the cutoff suggests that dropout risk becomes statistically detectable once eligibility is extended beyond roughly the top 6% of the within-school GPA distribution in PACE schools, corresponding to around the 60^{th} percentile of the nationwide distribution of high school test scores. This threshold should be interpreted as an approximate benchmark rather than an exact rule, but it highlights that mismatch may emerge as eligibility expands toward less-prepared students. However, preferential admission would continue to increase selective-college graduation well beyond this cutoff, suggesting that there is a trade-off for policymakers to evaluate.

Taken together, the findings shift the debate from *whether* preferential admission creates mismatch to *when* and *for whom* mismatch arises. They also highlight the value of rich longitudinal data and credible research designs for identifying meaningful tradeoffs within the targeted group. Expanding access to selective colleges can promote upward mobility for disadvantaged students who are sufficiently prepared to take advantage of the opportunity. But expanding eligibility without addressing academic preparation and information frictions can increase dropout and weaken the returns to college access. 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 promise of improving social mobility through affirmative action.

References

Akhtari, Mitra, Natalie Bau, and Jean-William Laliberté. 2024. "Affirmative action and precollege human capital." *American Economic Journal: Applied Economics* 16 (1):1–32.

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.

Antman, Francisca M, Brian Duncan, and Michael Lovenheim. 2024. "The long-run impacts of banning affirmative action in US higher education." *Oxford Review of Economic Policy* 40 (3):607–628.

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, Esteban M Aucejo, and V Joseph Hotz. 2016. "University differences in the graduation of minorities in STEM fields: Evidence from California." *American Economic Review* 106 (3):525–562.

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. 2025. "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely." Tech. rep.

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.

Bagde, Surendrakumar, Dennis Epple, and Lowell Taylor. 2016. "Does affirmative action work? Caste, gender, college quality, and academic success in India." *American Economic Review* 106 (6):1495–1521.

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.

Bertrand, Marianne, Rema Hanna, and Sendhil Mullainathan. 2010. "Affirmative action in education: Evidence from engineering college admissions in India." *Journal of Public Economics* 94 (1-2):16–29.

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.

———. 2023. "Affirmative action and its race-neutral alternatives." *Journal of Public Economics* 220:104839.

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.

Cingano, Federico, Filippo Palomba, Paolo Pinotti, and Enrico Rettore. 2025. "Making subsidies work: rules versus discretion." *Econometrica* 93 (3):747–778.

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.

Fort, Margherita, Andrea Ichino, Enrico Rettore, and Giulio Zanella. 2022. "Multi-cutoff rd designs with observations located at each cutoff: problems and solutions." .

Francis-Tan, Andrew and Maria Tannuri-Pianto. 2018. "Black Movement: Using discontinuities in admissions to study the effects of college quality and affirmative action." *Journal of Development Economics* 135:97–116.

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 and Salvador Navarro-Lozano. 2004. "Using matching, instrumental variables, and control functions to estimate economic choice models." *Review of Economics and statistics* 86 (1):30–57.

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.

Hinrichs, Peter. 2012. "The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities." *Review of Economics and Statistics* 94 (3):712–722.

———. 2014. "Affirmative action bans and college graduation rates." *Economics of Education Review* 42:43–52.

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.

LaLonde, Robert J. 1986. "Evaluating the econometric evaluations of training programs with experimental data." *The American economic review* :604–620.

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. 2025. "The direct and spillover effects of large-scale affirmative action at an elite Brazilian university." *Journal of Labor Economics* 43 (2):391–431.

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

Mello, Ursula. 2022. "Centralized admissions, affirmative action, and access of low-income students to higher education." *American Economic Journal: Economic Policy* 14 (3):166–197.

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

———. 2023b. "Sistema de Acceso y SUA desde 2016 100423." Report of the Chilean Ministry of Education.

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

Oliveira, Rodrigo, Alei Santos, and Edson Severnini. 2024. "Bridging the gap: Mismatch effects and catch-up dynamics under a Brazilian college affirmative action program." *Economics of Education Review* 98:102501.

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

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

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

Stinebrickner, Ralph and Todd R Stinebrickner. 2014. "A major in science? Initial beliefs and final outcomes for college major and dropout." *Review of Economic Studies* 81 (1):426–472.

Tincani, Michela M, Fabian Kosse, and Enrico Miglino. 2026. "Beliefs and the Incentive Effects of Preferential College Admissions: Evidence from an Experiment and a Structural Model." *CEPR DP18564* .

UCLA Higher Education Research Institute. 2023. "STEM Disciplines." https://www.heri.ucla.edu/PDFs/surveyAdmin/fac/Listing-of-STEM-Disciplines.pdf. [Online; accessed 11-May-2023].

UNESCO. 2011. "Campos De Educación 2013 De La Cine (ISCED-F 2013)." https://uis.unesco.org/sites/default/files/documents/isced-fields-of-education-and-training-2013-sp.pdf. [Online; accessed 11-May-2023].

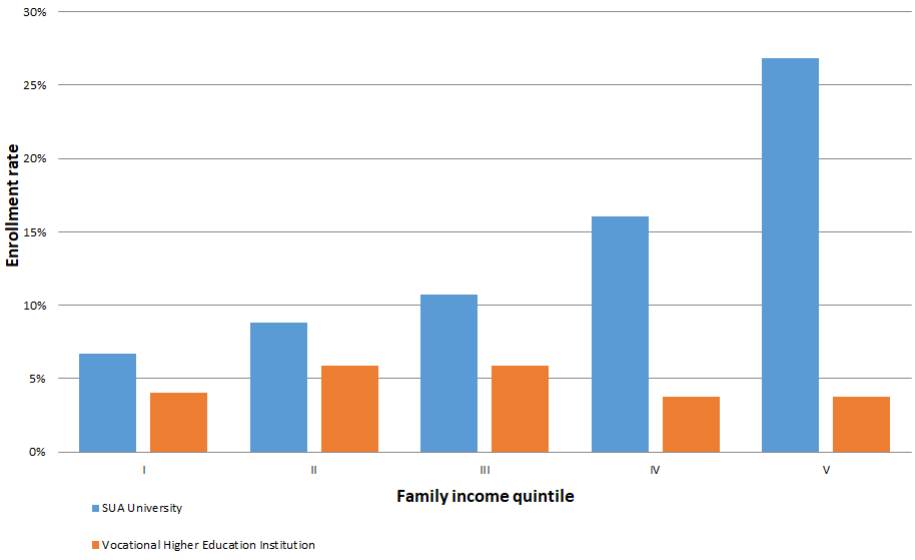
World Economic Forum. 2023. "Gender gap index in the area of economic participation and opportunity in Chile in 2023, by category." *Chart Statista* URL https://www.statista.com/statistics/803783/chile-gender-gap-labor-market-category/.

Appendix

A Additional Figures and Tables

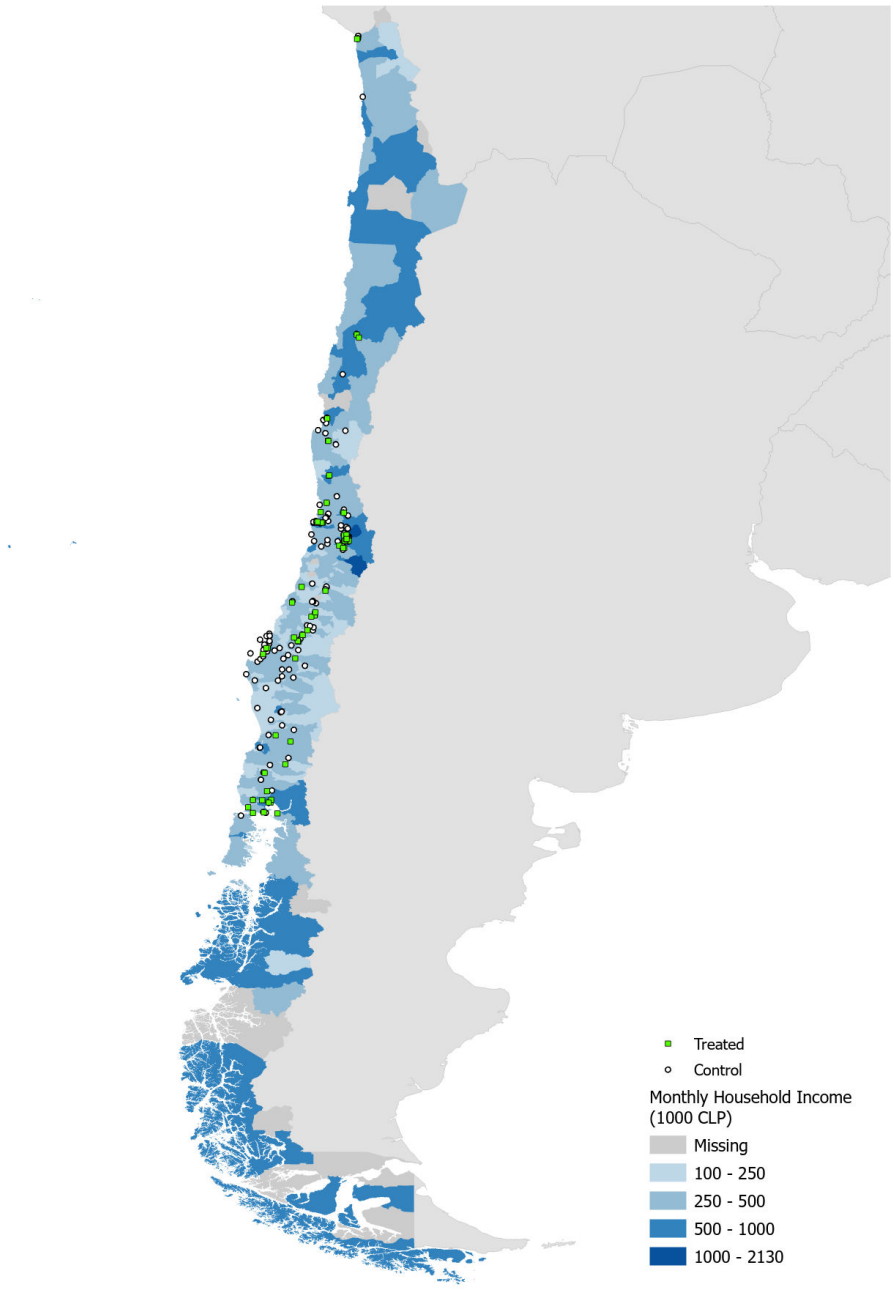
Figure A1: Graduation probability conditional on grade 10 standardized test score, by type of higher education institution.

Figure A2: Socioeconomic inequality in selective college enrollment.



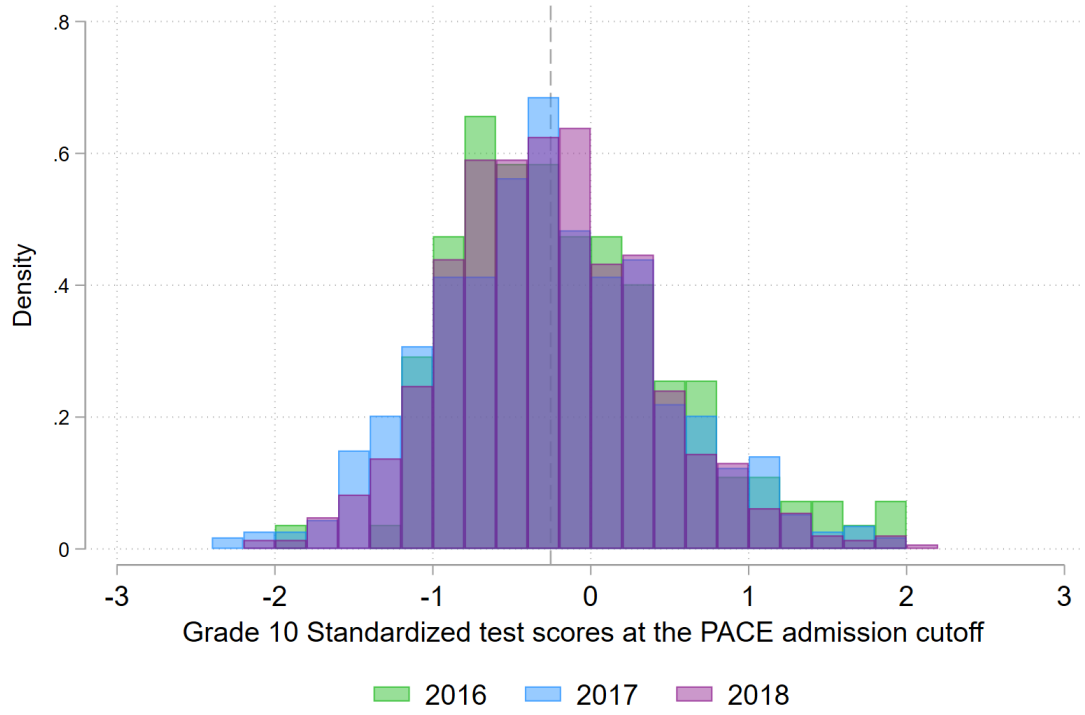
Notes: Percentage of 18-19 year old individuals who are enrolled in selective colleges and in vocational institutions in Chile, by family income quintile. Source: CASEN household survey, 2009, 2011 and 2013 waves.

Figure A3: Geographic distribution of schools in the experiment.



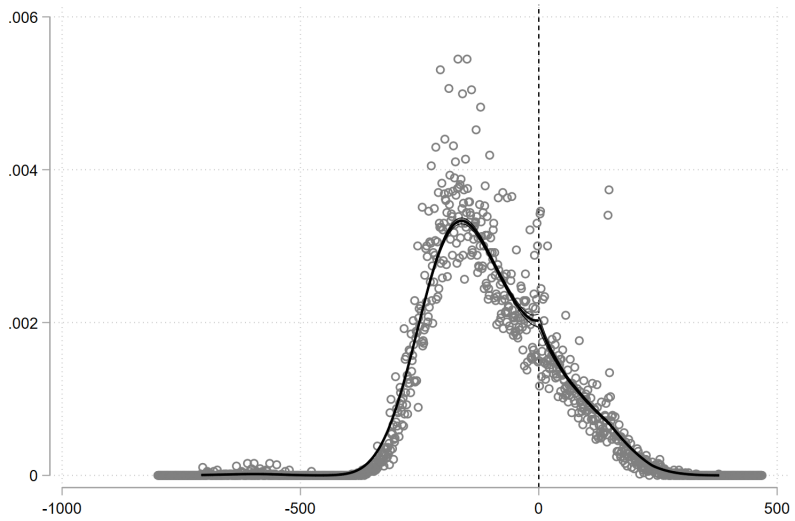
Notes: This figure shows the heatmap of Chile in terms of average monthly household income (1000 Chilean Pesos) and the geographic distribution of treated and control schools in the randomized experiment.

Figure A4: Academic preparation of PACE school students around the preferential admission cutoffs.



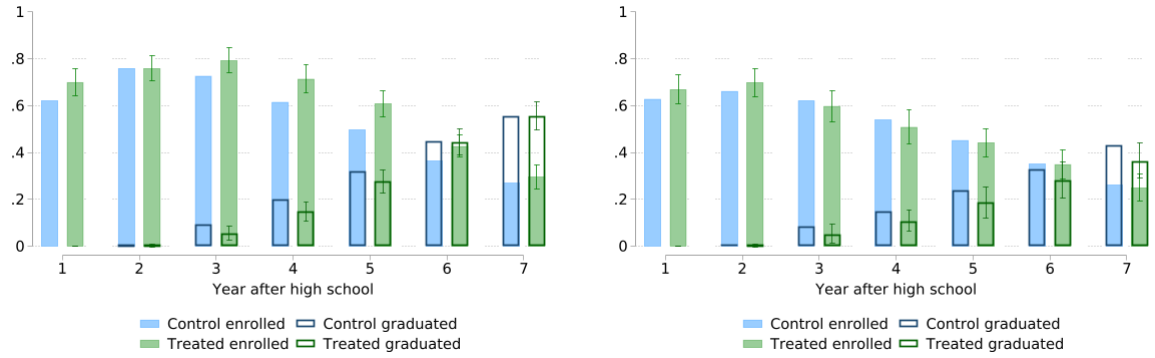
Notes: This figure shows the distribution of grade 10 SIMCE standardized test scores of students around the preferential admission cutoff. We obtained the SIMCE test score of the students whose school grades place them around the preferential admission cutoff in each school (within the optimal bandwidth 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 A5: 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 A6: Effects of PACE on higher education enrollment and graduation over time, RCT analysis *by gender*.

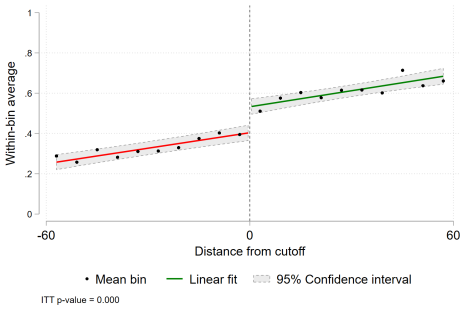


(a) Top 15% RCT sample, *Women*

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

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 or that have graduated from higher education, by gender. The error bars represent 95% confidence intervals for the difference in proportions across treatment groups (i.e., the treatment effect).

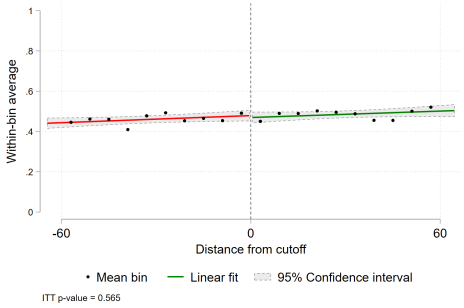
Figure A7: Intent to Treat Effects on Higher Education Outcomes.



(a) Ever enrolled in a selective college

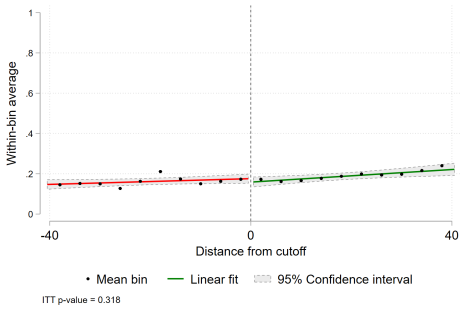
(b) Ever enrolled in any higher education institution

(c) Graduated from a selective college



(d) Graduated from any higher education institution

(e) Still enrolled in a selective college



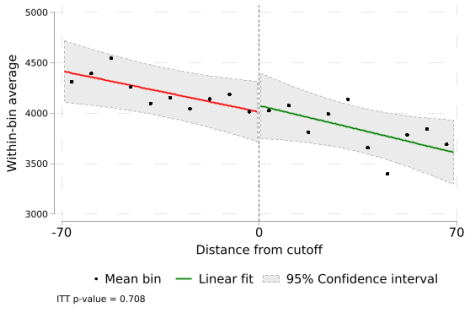
(f) Still enrolled in any higher education institution

(g) Dropped out from a selective college

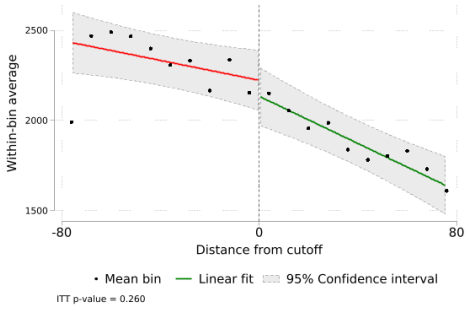
(h) Dropped out from any higher education institution

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 the school level. Outcomes are measured by year 7 after high school. Optimal bandwidth shown Calonico, Cattaneo, and Titiunik (2014).

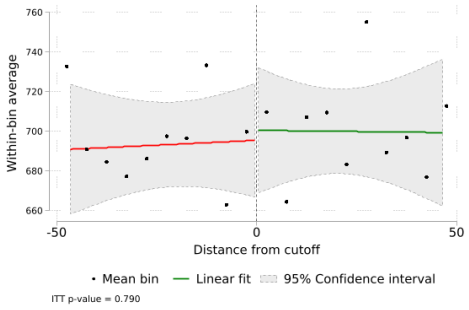
Figure A8: Intent to Treat Effects on Labor Market Outcomes.



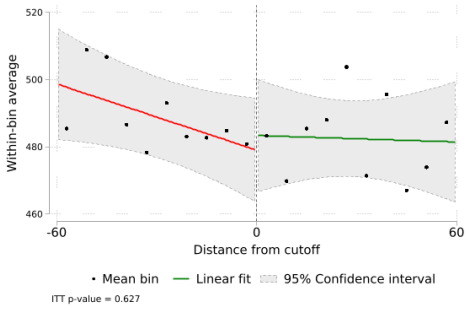
(a) Annual earnings in year 7



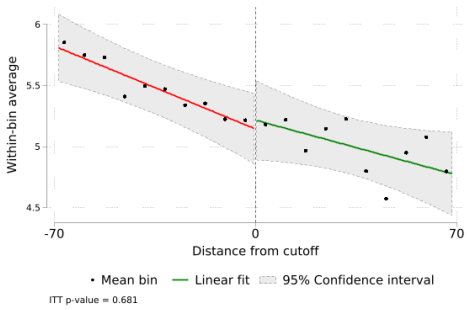
(b) Average annual earnings over years 1–7



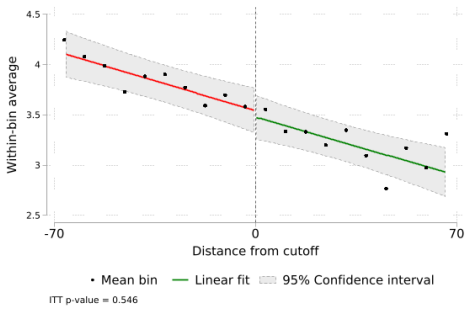
(c) Monthly income in year 7



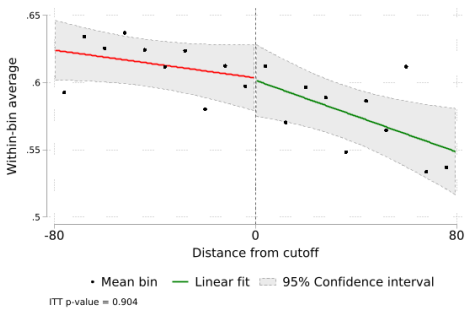
(d) Average monthly income over years 1–7



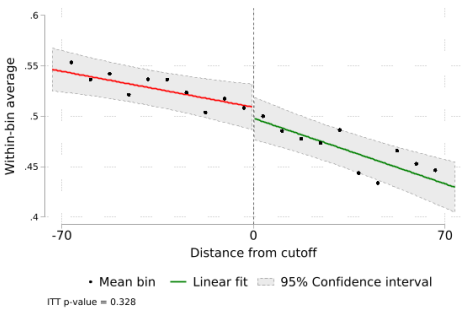
(e) Annual months worked in year 7



(f) Average annual months worked over years 1–7



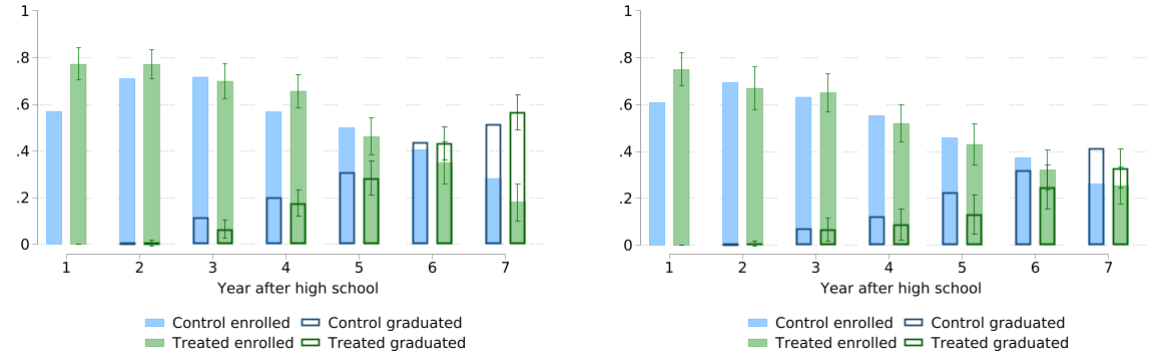
(g) Employed at least one month in year 7



(h) Average employment over years 1–7

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 the school level. Outcomes are measured after high school using labor-market administrative records. Optimal bandwidths are computed following Calonico, Cattaneo, and Titiunik (2014).

Figure A9: Effects of PACE on higher education enrollment and graduation over time, RD analysis.

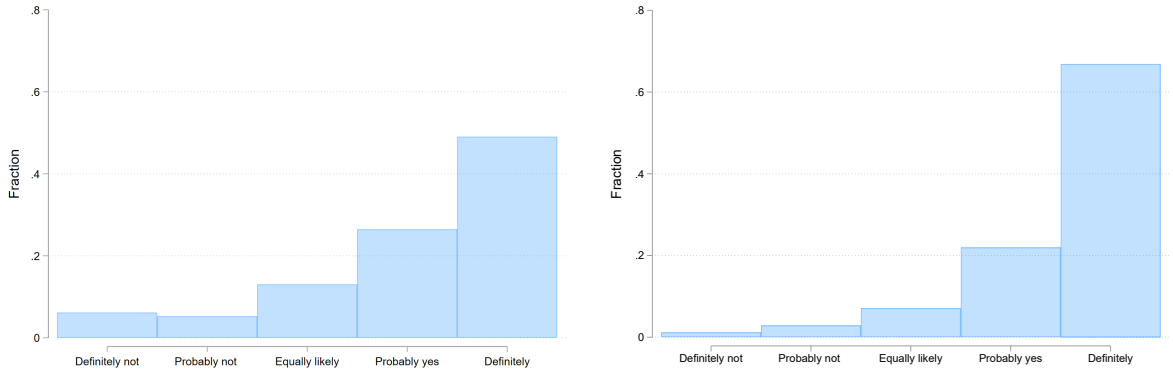


(a) RDD sample, *Women*

(b) RDD sample, *Men*

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). It represents δ as in equation 2. The x-axis indicates the year since high school graduation. The outcome variables are enrollment in higher education, and having graduated from higher education for each year. The bars represent 95% confidence intervals for δ calculated from standard errors clustered at school level. The regressions include the standard set of controls: gender, age, indicator for very-low-SES student, baseline SIMCE test score, never failed a grade, high school track (academic or vocational), and high school fixed effects. The control group fraction is the mean of the outcome just below the cutoff and within the optimal bandwidth, the treatment group fraction is obtained as the sum between the control group fraction and δ.

Figure A10: Beliefs about graduation chances.

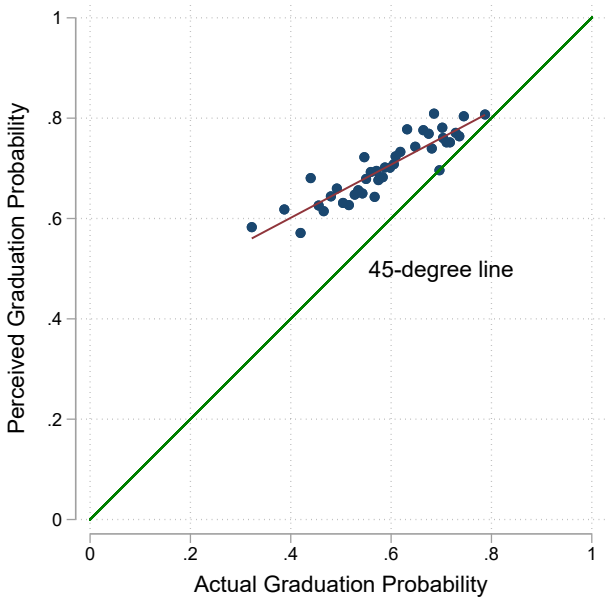


(a) All Students

(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 A11: Perceived and actual graduation chances: alternative coding



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 within seven years (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 seven 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.

Table A1: List of selective universities in 2018

Name	Type
UNIVERSIDAD DE VALPARAISO	State (CRUCH)
UNIVERSIDAD DE LA SERENA	State (CRUCH)
UNIVERSIDAD METROPOLITANA DE CIENCIAS DE LA EDUCACION	State (CRUCH)
UNIVERSIDAD ARTURO PRAT	State (CRUCH)
UNIVERSIDAD DE LA FRONTERA	State (CRUCH)
UNIVERSIDAD DE MAGALLANES	State (CRUCH)
UNIVERSIDAD DE O'HIGGINS	State (CRUCH)
UNIVERSIDAD TECNOLOGICA METROPOLITANA	State (CRUCH)
UNIVERSIDAD DE CHILE	State (CRUCH)
UNIVERSIDAD DE SANTIAGO DE CHILE	State (CRUCH)
UNIVERSIDAD DE ATACAMA	State (CRUCH)
UNIVERSIDAD DE ANTOFAGASTA	State (CRUCH)
UNIVERSIDAD DE LOS LAGOS	State (CRUCH)
UNIVERSIDAD DE PLAYA ANCHA DE CIENCIAS DE LA EDUCACION	State (CRUCH)
UNIVERSIDAD DEL BIO-BIO	State (CRUCH)
UNIVERSIDAD DE TALCA	State (CRUCH)
UNIVERSIDAD DE TARAPACA	State (CRUCH)
UNIVERSIDAD DE AYSEN	State (CRUCH)
PONTIFICIA UNIVERSIDAD CATOLICA DE VALPARAISO	Private (CRUCH)
UNIVERSIDAD CATOLICA DEL NORTE	Private (CRUCH)
UNIVERSIDAD ALBERTO HURTADO	Private (CRUCH)
UNIVERSIDAD AUSTRAL DE CHILE	Private (CRUCH)
UNIVERSIDAD CATOLICA DEL MAULE	Private (CRUCH)
UNIVERSIDAD DE CONCEPCION	Private (CRUCH)
UNIVERSIDAD TECNICA FEDERICO SANTA MARIA	Private (CRUCH)
PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE	Private (CRUCH)
UNIVERSIDAD CATOLICA DE LA SANTISIMA CONCEPCION	Private (CRUCH)
UNIVERSIDAD CATOLICA DE TEMUCO	Private (CRUCH)
UNIVERSIDAD DIEGO PORTALES	Private (CRUCH)
UNIVERSIDAD DE LOS ANDES	Private (CRUCH)
UNIVERSIDAD MAYOR	Private
UNIVERSIDAD CENTRAL DE CHILE	Private
UNIVERSIDAD ADOLFO IBAÑEZ	Private
UNIVERSIDAD CATOLICA CARDENAL RAUL SILVA HENRIQUEZ	Private
UNIVERSIDAD DEL DESARROLLO	Private
UNIVERSIDAD FINIS TERRAE	Private
UNIVERSIDAD AUTONOMA DE CHILE	Private
UNIVERSIDAD SAN SEBASTIAN	Private
UNIVERSIDAD ANDRES BELLO	Private

Notes: CRUCH universities are the traditional Chilean universities that belong to the 'Consejo de Rectores de las Universidades Chilenas'. Data provided by the Chilean Ministry of Education (Ministry of Education, 2023a).

Table A2: List of STEM majors

STEM majors
Biological Sciences
Physical Sciences
Natural Sciences, Mathematics and Statistics
Industry and Production
Engineering and Related Professions
Environmental Studies
Forestry, Agriculture, Fisheries
Health
Information and Communication Technology
Veterinary Medicine

Note.– The major categorization corresponds to the subarea categorization established by UNESCO in the CINE-F classification defined in 2013 and being used by the OECD since 2016 (UNESCO, 2011). Data provided by the Chilean Ministry of Education (Ministry of Education, 2023b). The distinction between STEM and Non-STEM majors follows the definition of STEM disciplines provided by the UCLA Higher Education Research Institute (2023).

Table A3: Summary statistics of main outcomes, RCT and RD samples

	Top 15%			RD		
	Obs.	Mean	S.D.	Obs.	Mean	S.D.
A. Higher education						
In any institution						
Ever enrolled	2437	0.87	0.34	13048	0.84	0.34
Graduation	2437	0.49	0.50	13048	0.41	0.50
Still enrolled	2437	0.18	0.38	13048	0.18	0.38
Dropout	2437	0.20	0.40	13048	0.26	0.42
In selective colleges						
Ever enrolled	2437	0.47	0.50	13048	0.44	0.50
Graduation	2437	0.21	0.40	13048	0.18	0.38
Still enrolled	2437	0.12	0.33	13048	0.11	0.31
Dropout	2437	0.09	0.28	13048	0.10	0.30
B. Labor market						
Avg earnings years 1-7	2437	2145.99	2454.81	13048	2176.26	2603.54
Avg earnings years 8-11	2437	5589.69	4377.84	13048	5400.52	4525.29
Avg earnings years 12-15	2437	6378.88	4429.50	13048	5937.09	7709.10
Avg months employed years 1-7	2437	3.56	3.39	13048	3.58	3.38
Avg months employed years 8-11	2437	5.87	3.61	13048	5.75	3.84
Avg months employed years 12-15	2437	5.29	2.76	13048	5.03	9.42
C. Higher education, heterogeneity by gender						
In any institution – Women						
Ever enrolled	1369	0.90	0.30	7730	0.88	0.33
Graduation	1369	0.55	0.50	7730	0.51	0.50
Still enrolled	1369	0.18	0.38	7730	0.17	0.37
Dropout	1369	0.17	0.37	7730	0.20	0.40
In any institution – Men						
Ever enrolled	1068	0.84	0.37	5318	0.84	0.36
Graduation	1068	0.41	0.49	5318	0.41	0.49
Still enrolled	1068	0.17	0.38	5318	0.18	0.38
Dropout	1068	0.25	0.44	5318	0.26	0.44
In selective colleges – Women						
Ever enrolled	1369	0.50	0.50	7730	0.45	0.50
Graduation	1369	0.25	0.43	7730	0.20	0.40
Still enrolled	1369	0.12	0.33	7730	0.10	0.31
Dropout	1369	0.08	0.27	7730	0.09	0.29
In selective colleges – Men						
Ever enrolled	1068	0.44	0.50	5318	0.43	0.50
Graduation	1068	0.15	0.36	5318	0.14	0.35
Still enrolled	1068	0.12	0.32	5318	0.11	0.32
Dropout	1068	0.10	0.31	5318	0.12	0.32
D. Labor market, heterogeneity by gender						
Women						
Avg earnings years 1-7	1369	1635.67	1975.35	7730	1746.44	2111.77
Avg earnings years 8-11	1369	4552.35	3772.03	7730	4414.24	3790.02
Avg earnings years 12-15	1369	4854.37	3522.99	7730	4514.15	3490.29
Avg months employed years 1-7	1369	2.98	3.06	7730	3.17	3.16
Avg months employed years 8-11	1369	5.11	3.41	7730	5.13	3.42
Avg months employed years 12-15	1369	4.42	2.42	7730	4.35	2.39
Men						
Avg earnings years 1-7	1068	2800.13	2827.12	5318	2801.03	3080.82
Avg earnings years 8-11	1068	6919.38	4728.96	5318	6834.13	5089.18
Avg earnings years 12-15	1068	8333.06	4698.63	5318	8005.42	10995.48
Avg months employed years 1-7	1068	4.30	3.64	5318	4.17	3.60
Avg months employed years 8-11	1068	6.84	3.62	5318	6.67	4.21
Avg months employed years 12-15	1068	6.41	2.76	5318	6.02	14.42

Notes: This table reports the number of observations, mean and standard deviations for the labor-market and higher-education outcomes for top 15% by the baseline high school GPA ranking of the experimental sample, and for the RD sample within the optimal bandwidth for the first-stage regression estimated using the method by Calonico, Cattaneo, and Titiunik (2014). Panels C and D show summary statistics by gender. Earnings are in 1000 Chilean pesos (December 2022 value). Earnings and months employed are coded as zero when the person is not observed in the Seguro de Cesantía dataset. Earnings and months employed in years 8-11 and 12-15 are predicted using the surrogate index (Athey et al., 2025). *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 A4: Summary statistics of baseline characteristics and main outcomes, predictions sample

	(1)	(2)	(3)
	Obs.	Mean	S.D.
A. Baseline characteristics			
Female	308899	0.52	0.50
Age	308899	16.44	1.27
Mother's education	48312	9.85	3.08
Father's education	47642	9.62	3.66
Family income	47842	187.41	169.43
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. Higher education outcomes			
In any institution			
Ever enrolled	308899	0.59	0.49
Graduation	308899	0.23	0.42
Still enrolled	308899	0.16	0.37
Dropout	308899	0.20	0.40
In selective colleges			
Ever enrolled	308899	0.14	0.35
Graduation	308899	0.04	0.21
Still enrolled	308899	0.07	0.26
Dropout	308899	0.03	0.17
C. Labor market outcomes			
Avg earnings years 1-7	308899	2355.43	2338.56
Avg earnings Years 8-11	308899	4737.88	4964.87
Avg earnings Years 12-15	308899	5593.91	6335.54
Avg months employed years 1-7	308899	4.79	3.72
Avg months employed Years 8-11	308899	6.20	4.80
Avg months employed Years 12-15	308899	5.84	4.85

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 graduated from high school in 2006-2010 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 graduated from high school in 2008 because only that cohort was surveyed. Earnings are in 1000 Chilean pesos (December 2022 value). Earnings and months employed are coded as zero when the person is not observed in the Seguro de Cesantía dataset. Labor market outcomes are observed in year 15 for the cohorts entering higher education in 2007-2009; in year 14 for the cohorts entering higher education in 2007-2010; in year 1-13 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 A5: Summary statistics SIMCE and PSU, experimental and RD samples

	Top 15%				RD			
	Obs.	Mean	Percentile	Percentile	Obs.	Mean	Percentile	Percentile
Summary Statistics				Black et al., 2023				Bleemer, 2021
SIMCE Score (Standardized)	2437	-0.01	0.52	0.86	13048	-0.21	0.46	.
PSU Score (Standardized)	2108	-0.26	0.42	.	11923	-0.48	0.34	0.72

Notes: This table reports the number of observations and mean for the SIMCE and PSU Score for the top 15% of the experimental sample 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). The table provides a comparison with summary statistics of comparable samples from Black, Denning, and Rothstein (2023) and Bleemer (2021). From Black, Denning, and Rothstein (2023) we use the following test scores: statewide percentile of score on 10th grade standardized tests (Texas high school exit exam, administered in 10th grade, see footnote 18 of Black et al. 2023). We show the average percentile among 'pulled-in students' from Black, Denning, and Rothstein (2023), who are defined as follows: students who pre-policy were unlikely to attend selective institutions and thanks to TTT are given access (i.e. those in top 10% of their school and who attend schools that pre-policy sent very few students to University of Texas, the most selective institution that TTT gives access to). Bleemer (2021) reports the average SAT score of California-resident freshman UC applicants: the table displays the nation-wide corresponding percentile for the whole sample (percentile of the SAT raw scores taken from the following source, year 2016: link).

Table A6: Sample Balance Across Treatment and Control Groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Age	Very Low SES	Mother education	Father education	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Female
A. Main results											
	Top 15%										
Treatment	0.049	-0.020	0.129	-0.019	5.756	0.084	-0.014	0.041	-0.013	0.075	0.001
	(0.051)	(0.025)	(0.177)	(0.229)	(12.545)	(0.121)	(0.018)	(0.066)	(0.019)	(0.072)	(0.055)
p-value	0.340	0.418	0.468	0.935	0.647	0.487	0.444	0.533	0.497	0.297	0.979
Mean	16.303	0.596	9.642	9.508	282.134	-0.041	0.941	0.155	0.043	0.281	0.561
S.d.	0.587	0.491	3.132	3.103	198.181	0.805	0.237	0.362	0.203	0.450	0.496
N	2437	2437	1914	1795	1919	2432	2437	2437	2437	2437	2437
B. Heterogeneity by gender											
	Top 15%										
Treatment - Women	0.055	-0.036	0.085	-0.015	0.616	0.088	-0.024	0.116	-0.009	0.105	
	(0.058)	(0.030)	(0.223)	(0.294)	(13.753)	(0.125)	(0.021)	(0.079)	(0.023)	(0.081)	
p-value	0.345	0.229	0.704	0.958	0.964	0.481	0.248	0.143	0.676	0.199	
Mean	16.296	0.621	9.568	9.502	267.251	-0.015	0.942	0.127	0.045	0.306	
S.d.	0.568	0.485	3.093	3.013	184.523	0.762	0.234	0.333	0.207	0.461	
N	1369	1369	1068	1000	1074	1368	1369	1369	1369	1369	
	Top 15%										
Treatment - Men	0.041	-0.000	0.189	-0.023	12.916	0.079	-0.001	-0.055	-0.018	0.037	
	(0.060)	(0.037)	(0.239)	(0.297)	(19.498)	(0.148)	(0.021)	(0.065)	(0.020)	(0.081)	
p-value	0.497	0.992	0.432	0.939	0.509	0.594	0.960	0.399	0.380	0.649	
Mean	16.311	0.563	9.734	9.515	300.883	-0.075	0.939	0.191	0.041	0.250	
S.d.	0.612	0.496	3.181	3.216	212.838	0.855	0.240	0.394	0.197	0.433	
N	1068	1068	846	795	845	1064	1068	1068	1068	1068	

Notes: In this table we regress pre-determined variables on treatment status among top 15% students according to the baseline high-school GPA ranking and gender. 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 A7: Tests for discontinuities in pre-determined variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Age	Very Low SES	Mother educ.	Father educ.	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Over-conf.	Female
						A. Main results						
Robust	0.078***	0.006	0.206	0.021	14.897	-0.023	-0.009	0.036	-0.003	-0.044	-0.080	0.009
	(0.028)	(0.024)	(0.173)	(0.187)	(16.514)	(0.058)	(0.011)	(0.039)	(0.017)	(0.043)	(0.130)	(0.029)
Bandwidth	61.498	72.583	69.298	72.238	81.019	74.528	92.260	67.623	83.743	65.930	58.535	66.636
Bandwidth N	12236	14430	9015	8853	10500	14680	18217	13448	16555	13048	562	13264
R-squared	0.005	0.000	0.001	0.001	0.000	0.041	0.008	0.001	0.000	0.005	0.003	0.000
Mean	16.263	0.610	9.990	9.732	305.467	-0.176	0.929	0.209	0.051	0.391	0.154	0.595
						B. Heterogeneity by gender						
Robust - Women	0.065**	-0.005	0.308	0.041	14.894	-0.036	-0.016	0.017	-0.002	-0.021	-0.124	
	(0.031)	(0.029)	(0.189)	(0.242)	(17.031)	(0.056)	(0.014)	(0.041)	(0.020)	(0.047)	(0.208)	
Bandwidth	93.904	71.230	88.591	60.534	70.239	74.707	92.255	83.424	73.912	72.716	63.032	
Bandwidth N	10884	8365	6720	4443	5376	8682	10724	9697	8663	8547	304	
R-squared	0.009	0.001	0.002	0.002	0.001	0.051	0.008	0.001	0.000	0.004	0.009	
Mean	16.267	0.640	9.773	9.521	285.232	-0.239	0.929	0.209	0.052	0.392	-0.092	
Robust - Men	0.042	0.005	-0.053	-0.130	50.452	-0.002	-0.002	0.017	0.003	-0.046	-0.167	
	(0.042)	(0.038)	(0.273)	(0.266)	(37.072)	(0.090)	(0.015)	(0.053)	(0.019)	(0.052)	(0.176)	
Bandwidth	70.198	63.761	75.875	94.825	45.616	70.169	90.174	79.704	67.749	77.367	79.453	
Bandwidth N	5690	5154	4011	4793	2479	5639	7317	6449	5486	6249	368	
R-squared	0.008	0.001	0.000	0.001	0.002	0.028	0.008	0.000	0.000	0.006	0.009	
Mean	16.291	0.568	10.184	10.131	328.310	-0.078	0.932	0.225	0.042	0.371	0.470	

Notes: In this table we report the estimate for coefficient ϕ in regression equation (3), by gender and using pre-determined 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 A8: ATE on earnings and monthly income (RCT analysis, top 15% sample) *with controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Earnings				Monthly income			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				A. Main results				
Treatment	-366.389	-252.306**	-179.981	-85.801	-13.531	-5.281	5.548	8.772
	(226.310)	(118.155)	(196.174)	(203.544)	(19.731)	(12.616)	(14.171)	(16.724)
Mean	4,155.396	2,200.768	5,605.954	6,369.485	704.708	498.610	767.474	994.603
Total obs.	2,437	2,437	2,437	2,437	1,454	1,982	2,437	2,437
				B. Heterogeneity by gender				
Treatment – Women	-155.293	-242.177*	52.291	191.255	-11.478	-3.677	26.283	43.110**
	(274.832)	(128.400)	(238.145)	(233.050)	(25.003)	(14.574)	(16.183)	(18.765)
Mean	3,254.633	1,679.663	4,472.070	4,740.071	642.087	449.189	691.511	888.710
Total obs.	1,369	1,369	1,369	1,369	744	1,084	1,369	1,369
Treatment – Men	-819.211**	-318.335	-675.864**	-643.564**	-29.190	-15.334	-31.870	-48.577**
	(340.080)	(200.422)	(284.906)	(300.163)	(26.527)	(20.678)	(19.995)	(22.384)
Mean	5,307.769	2,867.432	7,056.566	8,454.042	767.850	557.343	864.655	1,130.075
Total obs.	1,068	1,068	1,068	1,068	710	898	1,068	1,068
p-value Women=Men	0.097	0.728	0.029	0.011	0.605	0.623	0.009	0.000

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 include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school-region fixed effects. The standard errors reported in parentheses are clustered at the high school level. All panels show the observed impacts in year 7, the average impact between years 1 and 7 and the predicted average impacts between years 8 and 11 and between years 12 and 15 after high school graduation. In the first four columns, the dependent variable is the earnings in the year (or the average yearly earnings in the interval of years shown in the column heading). In the last four columns, the dependent variable is the average monthly earnings in the year conditional on working at least one month in that year (or the average of this variable in the interval of years shown in the column heading). Panel A shows impacts in the entire sample, while Panel B shows impacts in the sample of men and women separately. Earnings are measured in thousands of December 2022 Chilean pesos. Mean is the average of the outcome variable in the control group of the indicated sample. *** p<0.01, ** p<0.05, * p<0.10.

Table A9: ATE on months worked and employment (RCT analysis, top 15% sample) *with controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Months worked				Employment			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				A. Main results				
Treatment	-0.371	-0.368**	-0.178	-0.088	0.000	-0.022	-0.006	-0.010
	(0.231)	(0.162)	(0.158)	(0.128)	(0.021)	(0.017)	(0.014)	(0.012)
Mean	5.434	3.640	5.900	5.301	0.592	0.489	0.600	0.588
Total obs.	2,437	2,437	2,437	2,437	2,437	2,437	2,437	2,437
				B. Heterogeneity by gender				
Treatment – Women	-0.096	-0.371*	-0.009	0.028	0.015	-0.019	0.004	0.001
	(0.307)	(0.193)	(0.205)	(0.157)	(0.027)	(0.022)	(0.017)	(0.014)
Mean	4.620	3.054	5.064	4.374	0.529	0.441	0.530	0.519
Total obs.	1,369	1,369	1,369	1,369	1,369	1,369	1,369	1,369
Treatment – Men	-0.867**	-0.386	-0.522**	-0.330*	-0.036	-0.034	-0.031	-0.034*
	(0.351)	(0.261)	(0.237)	(0.195)	(0.032)	(0.027)	(0.022)	(0.017)
Mean	6.475	4.390	6.969	6.486	0.672	0.552	0.689	0.677
Total obs.	1,068	1,068	1,068	1,068	1,068	1,068	1,068	1,068
p-value Women=Men	0.078	0.961	0.078	0.117	0.198	0.646	0.180	0.086

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 include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school-region fixed effects. The standard errors reported in parentheses are clustered at the high school level. All panels show the observed impacts in year 7, the average impact between years 1 and 7 and the predicted average impacts between years 8 and 11 and between years 12 and 15 after high school graduation. In the first four columns, the dependent variable is number of months worked in the year (or the average yearly months worked in the interval of years shown in the column heading). In the last four columns, the dependent variable is a dummy for whether an individual is observed working at least one month in the year (or the average of this variable in the interval of years shown in the column heading). Panel A shows impacts in the entire sample, while Panel B shows impacts in the sample of men and women separately. Mean is the average of the outcome variable in the control group of the indicated sample. *** p<0.01, ** p<0.05, * p<0.10.

Table A10: Sample balance across treatment and control groups, conditional on being employed seven years after high school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Age	Very Low SES	Mother education	Father education	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Female
					A. Main results						
Treatment	0.064	0.006	0.144	-0.131	15.915	0.065	-0.018	0.020	-0.011	0.080	0.029
	(0.060)	(0.030)	(0.206)	(0.234)	(14.681)	(0.113)	(0.020)	(0.072)	(0.019)	(0.065)	(0.061)
p-value	0.285	0.837	0.486	0.575	0.280	0.568	0.374	0.784	0.572	0.225	0.638
Mean	16.315	0.602	9.499	9.433	274.444	-0.107	0.938	0.183	0.039	0.219	0.502
S.d.	0.582	0.490	3.140	3.018	182.292	0.792	0.241	0.387	0.194	0.414	0.500
N	1454	1454	1151	1081	1154	1450	1454	1454	1454	1454	1454
					B. Heterogeneity by gender						
Treatment - Women	0.034	-0.032	0.102	-0.213	5.434	0.083	-0.027	0.116	0.001	0.125	
	(0.068)	(0.040)	(0.277)	(0.323)	(16.411)	(0.123)	(0.023)	(0.087)	(0.024)	(0.078)	
p-value	0.622	0.421	0.713	0.510	0.741	0.497	0.248	0.184	0.950	0.109	
Mean	16.318	0.631	9.449	9.593	267.755	-0.056	0.934	0.151	0.037	0.249	
S.d.	0.573	0.483	3.153	2.868	184.435	0.741	0.249	0.358	0.189	0.433	
N	744	744	587	555	590	744	744	744	744	744	
Treatment - Men	0.099	0.046	0.196	-0.057	28.738	0.037	-0.007	-0.085	-0.024	0.025	
	(0.079)	(0.042)	(0.276)	(0.303)	(21.486)	(0.130)	(0.025)	(0.071)	(0.022)	(0.069)	
p-value	0.216	0.276	0.479	0.851	0.183	0.779	0.774	0.232	0.271	0.720	
Mean	16.312	0.574	9.549	9.270	281.152	-0.160	0.942	0.216	0.042	0.189	
S.d.	0.590	0.495	3.130	3.158	180.109	0.837	0.234	0.412	0.200	0.392	
N	710	710	564	526	564	706	710	710	710	710	

Notes: In this table we regress pre-determined variables on the treatment status of the top 15% students according to the baseline high-school GPA ranking and gender, conditional on being employed seven years after high school. 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 A11: ATE on education outcomes by higher-education institution *without controls* (RCT, baseline top 15%)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
				Main results				
Treatment	0.141***	0.050	0.038*	0.044***	0.020	-0.029	0.036*	0.012
	(0.044)	(0.033)	(0.020)	(0.014)	(0.020)	(0.027)	(0.020)	(0.020)
Total obs.	2437	2437	2437	2437	2437	2437	2437	2437
Mean	0.426	0.190	0.110	0.074	0.865	0.501	0.164	0.200

Notes: This table reports estimates of coefficient β from regression equation (1). The regressions include no controls. Standard errors clustered at the school level are shown in parentheses. *Mean* is the average value of the outcome variable in the control group. *** p<0.01, ** p<0.05, * p<0.10.

Table A12: ATE on earnings and monthly income (RCT analysis, top 15% sample) *without controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Earnings				Monthly income			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
	Main results							
Treatment	-232.527	-165.836	-44.648	22.283	-17.515	-5.549	15.275	17.929
	(308.511)	(194.909)	(289.441)	(332.438)	(22.837)	(15.385)	(19.565)	(26.485)
Mean	4,155.396	2,200.768	5,605.954	6,369.485	704.708	498.610	767.474	994.603
Total obs.	2,437	2,437	2,437	2,437	1,454	1,982	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 include no controls. Standard errors reported in parentheses are clustered at the high school level. Columns 3-4 and 7-8 use predicted outcomes from the surrogate-index pipeline. In the first four columns, the dependent variable is earnings in the year, or average yearly earnings in the interval of years shown in the column heading, assuming missing observations equal zero. In the last four columns, the dependent variable is average monthly earnings in the year conditional on working at least one month in that year, or the average of this variable in the interval of years shown in the column heading. Earnings are measured in thousands of December 2022 Chilean pesos. Mean is the average of the outcome variable in the control group. *** p<0.01, ** p<0.05, * p<0.10.

Table A13: ATE on months worked and employment (RCT analysis, top 15% sample) *without controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Months worked				Employment			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
	Main results							
Treatment	-0.233	-0.247	-0.078	-0.030	0.014	-0.010	0.003	-0.003
	(0.316)	(0.266)	(0.235)	(0.197)	(0.028)	(0.025)	(0.020)	(0.017)
Mean	5.434	3.640	5.900	5.301	0.592	0.489	0.600	0.588
Total obs.	2,437	2,437	2,437	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 include no controls. Standard errors reported in parentheses are clustered at the high school level. Columns 3-4 and 7-8 use predicted outcomes from the surrogate-index pipeline. In the first four columns, the dependent variable is number of months worked in the year, or the average yearly months worked in the interval of years shown in the column heading. In the last four columns, the dependent variable is an indicator for whether an individual is observed working at least one month in the year, or the average of this variable 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 A14: ATE on STEM and Non-STEM education outcomes by higher-education institution *with controls* (RCT, baseline top 15%)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
	Panel A: Main Results							
	STEM							
Treatment	0.094***	0.042**	0.026**	0.014	0.071***	0.023	0.022	0.013
	(0.022)	(0.017)	(0.013)	(0.010)	(0.025)	(0.021)	(0.014)	(0.013)
Mean	0.235	0.083	0.063	0.040	0.465	0.242	0.091	0.099
Total obs.	2437	2437	2437	2437	2437	2437	2437	2437
	Non-STEM							
Treatment	0.033	0.003	0.009	0.024**	-0.052*	-0.053**	0.008	-0.005
	(0.022)	(0.014)	(0.011)	(0.010)	(0.029)	(0.022)	(0.013)	(0.013)
Mean	0.221	0.107	0.048	0.039	0.474	0.263	0.077	0.114
Total obs.	2437	2437	2437	2437	2437	2437	2437	2437
	Panel B: Heterogeneity by gender							
	STEM							
Treatment – Women	0.110***	0.053**	0.044***	0.005	0.080**	0.033	0.024	0.011
	(0.025)	(0.021)	(0.015)	(0.010)	(0.035)	(0.028)	(0.017)	(0.015)
Mean Women	0.213	0.095	0.050	0.025	0.395	0.228	0.080	0.050
Total obs.	1369	1369	1369	1369	1369	1369	1369	1369
Treatment – Men	0.076**	0.031	0.000	0.027	0.037	-0.010	0.014	0.018
	(0.038)	(0.028)	(0.020)	(0.018)	(0.038)	(0.033)	(0.020)	(0.023)
Mean Men	0.264	0.068	0.080	0.059	0.554	0.260	0.106	0.161
Total obs.	1068	1068	1068	1068	1068	1068	1068	1068
p-value Women=Men	0.434	0.521	0.063	0.242	0.422	0.318	0.689	0.773
	Non-STEM							
Treatment – Women	0.063**	0.037*	0.010	0.025	-0.051	-0.032	0.006	-0.019
	(0.032)	(0.020)	(0.016)	(0.016)	(0.037)	(0.030)	(0.018)	(0.022)
Mean Women	0.257	0.127	0.058	0.038	0.567	0.330	0.092	0.120
Total obs.	1369	1369	1369	1369	1369	1369	1369	1369
Treatment – Men	-0.001	-0.034	0.006	0.022	-0.027	-0.061*	0.016	0.016
	(0.034)	(0.022)	(0.013)	(0.014)	(0.040)	(0.031)	(0.016)	(0.020)
Mean Men	0.175	0.081	0.036	0.039	0.355	0.179	0.057	0.106
Total obs.	1068	1068	1068	1068	1068	1068	1068	1068
p-value Women=Men	0.184	0.023	0.837	0.883	0.656	0.501	0.642	0.284

Notes: This table reports estimates of coefficient β from regression equation (1). Standard errors clustered at the school level are shown in parentheses. Panel A reports the main results for the full sample. Panel B reports results separately for women and men. At the bottom of Panel B, we report the p-value for the test of equality of treatment effects between women and men. Within each panel, results are reported separately for STEM and Non-STEM outcomes. *Mean* is the average value of the outcome variable in the control group for the relevant sample. The regressions include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school-region fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Selectivity		Ability distance		Rank	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
			A. Main results			
Treatment	0.176***	0.234***	0.072*	0.157***	-0.053***	-0.033**
	(0.036)	(0.034)	(0.041)	(0.041)	(0.017)	(0.015)
Total obs.	2437	2437	2437	2437	2437	2437
Selected obs.	2138	2138	2138	2138	2135	2135
Mean	0.025	0.025	0.010	0.010	0.459	0.459
			B: Heterogeneity by gender			
Treatment - Women	0.208***	0.319***	0.091*	0.235***	-0.085***	-0.040**
	(0.043)	(0.043)	(0.049)	(0.049)	(0.021)	(0.020)
Total obs.	1369	1369	1369	1369	1369	1369
Selected obs.	1238	1238	1238	1238	1237	1237
Mean	0.015	0.015	-0.003	-0.003	0.469	0.469
Treatment - Men	0.115**	0.138**	0.028	0.061	-0.026	-0.010
	(0.053)	(0.066)	(0.075)	(0.072)	(0.025)	(0.027)
Total obs.	1068	1068	1068	1068	1068	1068
Selected obs.	900	900	900	900	898	898
Mean	0.040	0.040	0.029	0.029	0.446	0.446

Notes: In this table we report the Lee bounds for the estimate of the coefficient β in regression equation (1). Appendix B.3 describes the methodology used to calculate bounds in the RDD. Panel A shows the estimates for the sample of students who belong to the top 15% of their school based on baseline GPA. Panel B displays the estimates for the same sample, divided by gender. 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: LATE on education outcomes by higher-education institution *without controls, with school fixed effects* (RD)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
				Main results				
Robust	0.187***	0.074***	-0.017	0.114***	0.061***	0.008	-0.040	0.090***
	(0.032)	(0.022)	(0.022)	(0.018)	(0.020)	(0.032)	(0.027)	(0.025)
Bandwidth	38	57	39	67	59	52	37	58
Bandwidth obs.	7752	11319	7904	13264	11728	10365	7684	11518
Mean	0.409	0.167	0.116	0.077	0.856	0.477	0.179	0.207

Notes: This table reports estimates of coefficient δ from regression equation (2). The regressions retain school fixed effects but exclude the standard control variables. 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, uniform kernels, and school fixed effects are used in all specifications. *** p<0.01, ** p<0.05, * p<0.10.

Table A17: LATE on earnings and monthly income (RD) *with controls*.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Earnings				Monthly income			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				A. Main results				
Robust	-7.497	-119.988	1.628	-52.332	-0.538	3.221	10.845	14.049
	(303.022)	(136.492)	(251.459)	(235.452)	(34.222)	(17.179)	(16.031)	(15.377)
Bandwidth	60.319	59.663	58.924	79.408	49.373	50.268	59.129	71.501
Bandwidth obs.	12,023	11,946	11,728	15,689	5,913	8,499	11,853	14,228
Mean	3,990.546	2,216.506	5,416.108	5,960.725	690.446	476.729	750.139	966.957
				B. Heterogeneity by gender				
Robust - Women	45.216	63.886	181.187	-13.044	23.286	7.218	31.849*	36.592**
	(358.765)	(161.363)	(284.626)	(249.952)	(39.147)	(20.117)	(18.060)	(16.295)
Bandwidth	41.850	38.030	48.387	49.394	40.917	41.001	51.180	72.026
Bandwidth obs.	5,060	4,666	5,769	5,870	2,783	4,171	6,096	8,473
Mean	3,309.085	1,745.204	4,429.620	4,615.524	608.914	427.509	670.167	859.971
Robust - Men	47.950	-69.310	13.832	-975.511*	-30.621	-19.748	-13.133	-25.757
	(484.980)	(248.840)	(425.038)	(584.545)	(56.974)	(27.981)	(29.061)	(31.063)
Bandwidth	61.153	57.066	61.973	48.616	48.239	44.975	59.586	61.537
Bandwidth obs.	4,972	4,651	4,999	3,980	2,573	3,133	4,856	4,999
Mean	4,881.648	2,659.922	6,792.572	7,846.068	786.080	545.809	859.576	1,127.908
p-value Women=Men	0.976	0.634	0.698	0.224	0.566	0.430	0.203	0.099
				C. Heterogeneity by overconfidence				
Robust - High	-1685.962	88.233	-1594.576	-1797.419	-56.497	-51.505	-114.621	-56.408
	(1658.042)	(1102.659)	(1539.557)	(1326.278)	(208.312)	(118.505)	(122.685)	(100.260)
Bandwidth	80.909	64.163	60.297	55.670	66.872	56.545	54.151	72.815
Bandwidth obs.	449	358	342	321	236	280	314	412
Mean	3,658.048	1,979.383	5,237.132	6,557.639	616.614	472.742	763.645	997.209
Robust - Low	-3838.268	-1556.063	-2891.946	-1667.726	-106.464	-27.412	-46.655	24.001
	(2568.044)	(1043.626)	(1962.302)	(2464.514)	(162.375)	(75.401)	(103.175)	(151.659)
Bandwidth	51.205	52.619	49.643	43.646	72.489	66.579	61.143	51.028
Bandwidth obs.	201	208	197	174	168	219	238	201
Mean	3,277.494	2,278.832	4,376.410	4,754.346	629.891	435.953	659.625	844.235
p-value High=Low	0.311	0.310	0.413	0.845	0.911	0.962	0.747	0.806

Notes: This table shows the estimate of parameter δ in equation (2). The regressions include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade and high school track (academic or vocational). Panel A and Panel B additionally include school fixed effects; Panel C excludes school fixed effects. The standard errors reported in parentheses are clustered at the high school level. All panels show the observed impacts in year 7, the average impact between years 1 and 7 and the predicted average impacts between years 8 and 11 and between years 12 and 15 after high school graduation. In the first four columns, the dependent variable is the earnings in the year (or the average yearly earnings in the interval of years shown in the column heading). In the last four columns, the dependent variable is the average monthly earnings in the year (or the average of this variable in the interval of years shown in the column heading). Panel A shows impacts in the entire sample, Panel B shows impacts in the sample of men and women separately, and Panel C shows impacts in the subsamples of above median and below median overconfidence separately. 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: Bounds on selectivity and rank effects in higher education course (RD, all students)

	(1)	(2)	(3)	(4)	(5)	(6)
	Selectivity		Ability distance		Rank	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
			A. Main results			
Robust	0.128***	0.262***	0.055	0.330***	-0.121***	-0.029
	(0.037)	(0.035)	(0.050)	(0.048)	(0.019)	(0.022)
Bandwidth	58	58	58	58	59	59
Observations	9761	9762	9761	9762	9940	9941
Mean	-0.062	-0.062	0.054	0.054	0.445	0.445
			B. Heterogeneity by gender			
Robust - Women	0.108**	0.268***	0.022	0.323***	-0.124***	-0.020
	(0.051)	(0.047)	(0.069)	(0.065)	(0.027)	(0.030)
Bandwidth	50	50	50	50	48	48
Observations	5089	5090	5089	5090	4883	4884
Mean	-0.100	-0.100	0.064	0.064	0.442	0.442
Robust - Men	0.156***	0.201***	0.134*	0.236***	-0.104***	-0.053*
	(0.055)	(0.054)	(0.074)	(0.073)	(0.028)	(0.031)
Bandwidth	70	70	70	70	73	73
Observations	4791	4792	4791	4792	4906	4907
Mean	-0.001	-0.001	0.029	0.029	0.446	0.446

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2). Appendix B.3 describes the methodology used to calculate bounds in the RDD. Panel A shows the estimates for all students who belong to the RD sample of students. Panel B displays the estimates for the same sample divided by gender. 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 A19: LATE on months worked and employment (RD) *with controls.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Months worked				Employment			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				A. Main results				
Robust	-0.027	-0.193	-0.081	0.043	-0.015	-0.031*	-0.005	0.012
	(0.298)	(0.176)	(0.204)	(0.159)	(0.028)	(0.018)	(0.018)	(0.016)
Bandwidth	65.075	65.996	64.401	106.718	64.364	64.850	68.905	82.652
Bandwidth obs.	12,969	13,048	12,784	21,158	12,784	12,872	13,668	16,371
Mean	5.108	3.526	5.731	4.969	0.601	0.511	0.599	0.572
				B. Heterogeneity by gender				
Robust - Women	-0.261	-0.312	-0.236	-0.187	-0.041	-0.050**	-0.015	-0.026
	(0.394)	(0.216)	(0.255)	(0.203)	(0.037)	(0.022)	(0.022)	(0.019)
Bandwidth	58.247	55.542	56.250	42.047	51.324	58.345	52.403	40.692
Bandwidth obs.	6,910	6,579	6,657	5,114	6,096	6,910	6,218	4,959
Mean	4.928	3.287	5.182	4.467	0.579	0.485	0.553	0.541
Robust - Men	0.446	0.051	0.324	-0.504	0.054	-0.009	0.071**	-0.019
	(0.435)	(0.298)	(0.306)	(0.552)	(0.048)	(0.029)	(0.032)	(0.047)
Bandwidth	61.141	57.811	59.199	54.332	43.605	64.123	51.247	54.798
Bandwidth obs.	4,972	4,687	4,814	4,415	3,579	5,199	4,184	4,451
Mean	5.476	3.975	6.384	5.771	0.628	0.536	0.662	0.620
p-value Women=Men	0.329	0.533	0.251	0.641	0.218	0.439	0.083	0.958
				C. Heterogeneity by overconfidence				
Robust - High	-2.025	0.340	-1.494	-1.602**	-0.166	-0.038	-0.134	-0.096
	(1.554)	(1.115)	(1.066)	(0.792)	(0.163)	(0.123)	(0.094)	(0.064)
Bandwidth	76.609	50.405	57.079	53.596	62.027	50.273	59.914	71.874
Bandwidth obs.	430	290	327	312	353	290	340	408
Mean	5.427	3.393	6.022	5.762	0.644	0.537	0.641	0.639
Robust - Low	-1.762	-2.921*	-1.540	-0.669	-0.422	-0.214	-0.176	-0.097
	(2.071)	(1.535)	(1.456)	(1.298)	(0.266)	(0.155)	(0.137)	(0.124)
Bandwidth	64.992	58.972	61.962	52.822	56.380	61.015	57.718	52.605
Bandwidth obs.	246	229	241	208	221	238	227	208
Mean	4.397	4.224	5.406	4.572	0.655	0.559	0.578	0.551
p-value High=Low	0.934	0.173	0.800	0.712	0.272	0.527	0.717	0.940

Notes: This table shows the estimate of parameter δ in equation (2). The regressions include the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade and high school track (academic or vocational). Panel A and Panel B additionally include school fixed effects; Panel C excludes school fixed effects. The standard errors reported in parentheses are clustered at the high school level. All panels show the observed impacts in year 7, the average impact between years 1 and 7 and the predicted average impacts between years 8 and 11 and between years 12 and 15 after high school graduation. In the first four columns, the dependent variable is number of months worked in the year (or the average yearly months worked in the interval of years shown in the column heading). In the last four columns, the dependent variable is a dummy for whether an individual is observed working at least one month in the year (or the average of this variable in the interval of years shown in the column heading). Panel A shows impacts in the entire sample, Panel B shows impacts in the sample of men and women separately, and Panel C shows impacts in the subsamples of above median and below median overconfidence separately. 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: Tests for discontinuities in pre-determined variables, conditional on being employed seven years after high school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Age	Very Low SES	Mother educ.	Father educ.	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Over-conf.	Female
						A. Main results						
Robust	0.052	-0.015	0.234	0.131	28.889	-0.005	0.009	-0.006	0.010	-0.007	0.109	-0.031
	(0.041)	(0.032)	(0.242)	(0.261)	(19.788)	(0.067)	(0.018)	(0.046)	(0.017)	(0.044)	(0.228)	(0.038)
Bandwidth	69.385	69.385	69.385	69.385	69.385	69.385	69.385	69.385	69.385	69.385	69.385	69.385
Bandwidth N	8300	8300	5383	5095	5384	8218	8300	8300	8300	8300	406	8300
R-squared	0.007	0.000	0.001	0.000	0.001	0.027	0.005	0.001	0.000	0.006	0.022	0.002
Mean	16.284	0.608	9.831	9.569	292.711	-0.224	0.914	0.235	0.046	0.326	0.204	0.579
						B. Heterogeneity by gender						
Robust - Women	-0.003	-0.014	0.043	0.173	41.573	-0.080	0.019	0.042	-0.016	-0.007	-0.546	
	(0.069)	(0.058)	(0.436)	(0.491)	(32.231)	(0.101)	(0.035)	(0.065)	(0.024)	(0.068)	(0.452)	
Bandwidth	35.070	35.070	35.070	35.070	35.070	35.070	35.070	35.070	35.070	35.070	35.070	
Bandwidth N	2440	2440	1582	1497	1580	2417	2440	2440	2440	2440	111	
R-squared	0.000	0.002	0.001	0.002	0.002	0.010	0.002	0.001	0.001	0.002	0.026	
Mean	16.288	0.666	9.722	9.236	258.424	-0.243	0.915	0.243	0.047	0.351	-0.122	
Robust - Men	0.100	0.017	0.752**	0.383	56.298	0.049	0.012	-0.054	0.024	-0.001	-0.097	
	(0.070)	(0.053)	(0.370)	(0.423)	(35.048)	(0.100)	(0.030)	(0.064)	(0.020)	(0.057)	(0.336)	
Bandwidth	60.232	60.232	60.232	60.232	60.232	60.232	60.232	60.232	60.232	60.232	60.232	
Bandwidth N	3226	3226	2114	2018	2116	3195	3226	3226	3226	3226	184	
R-squared	0.007	0.001	0.001	0.001	0.002	0.017	0.004	0.000	0.000	0.005	0.010	
Mean	16.293	0.579	9.896	9.749	325.307	-0.154	0.910	0.261	0.037	0.306	0.484	

Notes: In this table we report the estimate for coefficient ϕ in regression equation (3), by gender and using pre-determined variables as the dependent variable. Standard errors clustered at the school level are shown in parentheses. The sample is restricted to students that are employed seven years after high school. *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 A21: LATE on months worked and employment (RD) *without controls, with school fixed effects.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Months worked				Employment			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				Main results				
Robust	-0.014	-0.166	0.015	0.061	-0.004	-0.028	-0.001	0.013
	(0.284)	(0.169)	(0.201)	(0.209)	(0.028)	(0.019)	(0.017)	(0.016)
Bandwidth	72.912	70.905	68.705	87.655	67.239	63.981	69.733	68.397
Bandwidth obs.	14,430	14,029	13,668	17,363	13,349	12,682	13,838	13,548
Mean	5.108	3.526	5.731	4.969	0.601	0.511	0.599	0.572

Notes: This table shows the estimate of parameter δ in equation (2). The regressions retain school fixed effects but exclude the standard control variables. Standard errors reported in parentheses are clustered at the high school level. Columns 3-4 and 7-8 use predicted outcomes from the surrogate-index pipeline. In the first four columns, the dependent variable is number of months worked in the year, or the average yearly months worked in the interval of years shown in the column heading. In the last four columns, the dependent variable is an indicator for whether an individual is observed working at least one month in the year, or the average of this variable 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 bandwidths, a linear polynomial of the ranking score, uniform kernels, and school fixed effects are used in all specifications. *** p<0.01, ** p<0.05, * p<0.10.

Table A22: LATE on earnings and monthly income (RD) *without controls, with school fixed effects.*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Earnings with zero				Monthly income			
	Year 7	Years 1-7	Years 8-11	Years 12-15	Year 7	Years 1-7	Years 8-11	Years 12-15
				Main results				
Robust	3.426	-105.552	-92.675	-35.417	13.214	2.804	3.617	-2.315
	(301.402)	(138.859)	(249.041)	(230.970)	(33.135)	(16.961)	(16.295)	(17.412)
Bandwidth	61.015	61.727	63.874	69.795	51.669	53.825	64.131	59.059
Bandwidth obs.	12,236	12,313	12,682	13,838	6,212	9,042	12,784	11,853
Mean	3,990.546	2,216.506	5,416.108	5,960.725	690.446	476.729	750.139	966.957

Notes: This table shows the estimate of parameter δ in equation (2). The regressions retain school fixed effects but exclude the standard control variables. Standard errors reported in parentheses are clustered at the high school level. Columns 3-4 and 7-8 use predicted outcomes from the surrogate-index pipeline. In the first four columns, the dependent variable is earnings in the year, or average yearly earnings in the interval of years shown in the column heading, assuming missing observations equal zero. In the last four columns, the dependent variable is average monthly earnings in the year conditional on working at least one month in that year, or the average of this variable in the interval of years shown in the column heading. Earnings are measured in thousands of December 2022 Chilean pesos. 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 bandwidths, a linear polynomial of the ranking score, uniform kernels, and school fixed effects are used in all specifications. *** p<0.01, ** p<0.05, * p<0.10.

Table A23: LATE on STEM and Non-STEM education outcomes by higher-education institution *with controls* (RD)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Selective College				Any Institution			
	Ever enrolled	Graduation	Still enrolled	Dropout	Ever enrolled	Graduation	Still enrolled	Dropout
	Panel A: Main Results							
	STEM							
Robust	0.036	-0.004	-0.005	0.043***	0.001	-0.011	0.013	0.004
	(0.030)	(0.014)	(0.016)	(0.014)	(0.034)	(0.022)	(0.016)	(0.016)
Bandwidth	33	52	53	56	45	68	59	74
Bandwidth obs.	6869	10473	10645	11099	9186	13448	11853	14731
Mean	0.216	0.065	0.056	0.040	0.430	0.211	0.080	0.106
	Non-STEM							
Robust	0.173***	0.085***	-0.013	0.092***	0.076***	0.012	-0.030	0.088***
	(0.027)	(0.020)	(0.015)	(0.017)	(0.026)	(0.027)	(0.022)	(0.023)
Bandwidth	58	50	42	45	68	55	32	39
Bandwidth obs.	11518	9990	8584	9186	13548	11003	6706	7904
Mean	0.221	0.103	0.057	0.042	0.496	0.276	0.097	0.112
	Panel B: Heterogeneity by gender							
	STEM							
Robust - Women	0.004	0.014	-0.018	0.017	0.007	0.041	-0.005	-0.013
	(0.036)	(0.021)	(0.021)	(0.015)	(0.043)	(0.034)	(0.023)	(0.018)
Bandwidth	36	47	46	54	42	47	44	62
Bandwidth obs.	4384	5606	5511	6365	5060	5661	5274	7362
Mean	0.199	0.063	0.051	0.033	0.357	0.176	0.068	0.082
Robust - Men	0.018	-0.045*	0.021	0.078***	-0.086	-0.108***	0.004	0.052
	(0.044)	(0.027)	(0.022)	(0.025)	(0.054)	(0.040)	(0.025)	(0.033)
Bandwidth	39	42	61	58	42	60	74	63
Bandwidth obs.	3177	3416	4920	4687	3416	4856	5981	5120
Mean	0.229	0.072	0.066	0.053	0.533	0.247	0.104	0.150
p-value Women=Men	0.419	0.226	0.174	0.034	0.317	0.013	0.998	0.347
	Non-STEM							
Robust - Women	0.187***	0.090***	-0.023	0.123***	0.074*	0.016	-0.082***	0.101***
	(0.032)	(0.026)	(0.017)	(0.023)	(0.039)	(0.037)	(0.027)	(0.029)
Bandwidth	62	54	56	39	49	52	36	43
Bandwidth obs.	7314	6419	6579	4727	5870	6155	4452	5233
Mean	0.239	0.112	0.061	0.032	0.550	0.334	0.110	0.112
Robust - Men	0.169***	0.061**	0.018	0.061***	0.080*	0.008	0.019	0.053*
	(0.036)	(0.027)	(0.025)	(0.023)	(0.043)	(0.035)	(0.031)	(0.031)
Bandwidth	66	55	45	58	62	64	44	56
Bandwidth obs.	5318	4486	3720	4687	5041	5154	3628	4520
Mean	0.197	0.085	0.050	0.055	0.389	0.180	0.085	0.115
p-value Women=Men	0.849	0.801	0.185	0.191	0.383	0.673	0.028	0.633

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. Panel A reports the main results for the full sample. Panel B reports results separately for women and men. At the bottom of Panel B, we report the p-value for the test of equality of treatment effects between women and men. Within each panel, results are reported separately for STEM and Non-STEM outcomes. *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 standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school fixed effects. *** p<0.01, ** p<0.05, * p<0.10.

Table A24: Balance Across Samples with and without Missing Overconfidence Data

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Age	Very Low SES	Mother educ.	Father educ.	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Female
Missing	-0.036	0.027	-0.129	-0.058	6.741	-0.004	0.012	0.056	0.022	0.019	0.087
	(0.039)	(0.022)	(0.177)	(0.253)	(14.229)	(0.110)	(0.014)	(0.057)	(0.018)	(0.066)	(0.055)
p-value	0.361	0.227	0.465	0.820	0.636	0.973	0.368	0.330	0.210	0.775	0.114
Mean	16.365	0.586	10.101	9.848	303.788	-0.204	0.911	0.166	0.029	0.351	0.510
S.d.	0.644	0.493	3.013	3.326	230.842	0.811	0.285	0.372	0.169	0.478	0.500
N	13048	13048	8565	8090	8554	12929	13048	13048	13048	13048	13048

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 A25: Perceived and actual chances of graduation from selective college among those who enroll.

Survey Answer	**Actual % Graduates**
	(1)
Chances of Graduating $\leq 50\%$	27.67
Will Probably Graduate	39.62
Will Certainly Graduate	41.76
Any Survey Answer	39.72

Notes: The table uses the sample of students who enrolled in a selective college during the 7 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 seven years after leaving high school (2024). Each row restricts the sample according to students' survey answers, and shows among the students who gave each answer what percentage have graduated from a selective college seven years later.

Table A26: Gender gap in overconfidence

	(1)	(2)
	Overconfidence	Overconfidence
Female	-0.289***	-0.284***
	(0.032)	(0.032)
Controls	NO	YES
Obs.	5770	5770
Mean	0.134	0.134

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 B.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 A27: Bounds on selectivity and rank effects in higher education course (RD) by overconfidence

	(1)	(2)	(3)	(4)	(5)	(6)
	Selectivity		Ability distance		Rank	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
	A. Heterogeneity by overconfidence					
Robust - High	0.156	0.505**	0.751*	1.287***	-0.283*	-0.143
	(0.281)	(0.237)	(0.430)	(0.412)	(0.167)	(0.162)
Bandwidth	65	65	65	65	65	65
Observations	314	313	314	313	314	313
Mean	-0.025	-0.025	0.035	0.035	0.446	0.446
Robust - Low	0.238	0.346	0.253	0.284	-0.039	-0.009
	(0.308)	(0.303)	(0.448)	(0.427)	(0.161)	(0.157)
Bandwidth	83	83	83	83	83	83
Observations	261	260	261	260	261	260
Mean	-0.154	-0.154	0.063	0.063	0.454	0.454

Notes: In this table we report the bounds on the estimate for coefficient δ in regression equation (2) for the subsamples of students with above median and below median overconfidence. School fixed effects are not included. 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: ATE on high school GPA rank (RCT, all students)

	(1)	(2)
	GPA rank	GPA rank
Treatment	-0.001	-0.015
	(0.001)	(0.016)
Total obs.	13393	13393
Mean	0.517	0.517

Notes: This table reports estimates of coefficient β from regression equation (1), estimated on the full sample of students in experimental schools. Standard errors clustered at the school level are shown in parentheses. Column 1 includes no controls, column 2 includes the standard set of controls: grade 10 standardized test score (SIMCE), gender, age, an indicator for very low socioeconomic status (*Alumno Prioritario*), an indicator for never having repeated a grade, high school track (academic or vocational), and school-region fixed effects. *Mean* is the average value of the outcome variable in the control group. The outcome variable is the within-school rank in terms of GPA in all four high school years. *** p<0.01, ** p<0.05, * p<0.10.

Table A29: Sample Balance Across Treatment and Control Groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Age	Very Low SES	Mother education	Father education	Family income	SIMCE score	Never failed	Santiago resident	Rural school	Academic track	Female
Treatment	0.021	0.011	0.110	0.116	15.153	0.076	-0.014	0.040	-0.011	0.030	-0.016
	(0.039)	(0.017)	(0.153)	(0.166)	(10.673)	(0.092)	(0.017)	(0.065)	(0.017)	(0.063)	(0.048)
p-value	0.588	0.533	0.471	0.484	0.157	0.412	0.418	0.538	0.524	0.635	0.736
Mean	16.550	0.605	9.524	9.319	283.132	-0.596	0.839	0.151	0.038	0.254	0.493
S.d.	0.772	0.489	3.121	3.204	203.917	0.747	0.368	0.358	0.191	0.435	0.500
N	14414	14414	9668	9157	9701	14307	14414	14414	14414	14414	14414

Notes: In this table we regress pre-determined variables on the treatment status of all experimental students according to the baseline high-school GPA ranking and gender. 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.

B Data Appendix

B.1 Variable construction

Labor force participation. We measure employment using monthly records from the social security registry (*Seguro de Cesantía*, SC). *Months worked in year 7* is the number of months in year 7 after high school graduation in which the individual is observed employed in the registry; it is zero when no employment month is observed. *Months worked in years 1-7* is the average number of months worked each year, in the period from 1 to 7 years post high school graduation. *Employment in year 7* is an indicator equal to one if the individual is observed employed for at least one month in year 7, and zero otherwise. *Employment in years 1-7* is the average of this yearly employment indicator over years 1 through 7 post high school graduation.

Earnings. As in other studies using the SC dataset (Neilson et al., 2021), we collapse earnings observations for an individual to the annual level. Earnings are constructed from monthly employment records in the social security registry. Monthly earnings are deflated to December 2022 pesos using the CPI, divided by 1,000, summed across employers within month.[40] Therefore, they are measured in thousands of December 2022 Chilean Pesos (exchange rate: CLP $1,000,000 \sim$ USD $1,000$). *Earnings in year 7* is total annual earnings in year 7 post high school, with missing monthly or annual earnings replaced by zero. *Earnings in years 1-7* is the average of annual earnings in the years 1 to 7, representing average annual earnings over this period, with missing monthly or annual earnings replaced by zero. *Monthly income in year 7* is annual earnings in year 7 divided by the number of months observed employed in that year, and is missing when the individual is not observed employed that year. Therefore, it represents average monthly income when working. *Monthly income in years 1-7* is the average of monthly incomes from year 1 to year 7 over non-missing years.

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 share within the year. We use the following categorization of sectors:

- Agriculture, livestock, fishing
- Mining
- Manufacturing
- Electricity, gas

[40] Further, we trim monthly earnings observations above the calendar-year 99th percentile in our sample.

- 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 institutions 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, 2023a). 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.

We define the following outcomes for year 7 after high school graduation. For type $k \in \{\text{Selective College}, \text{Any Institution}\}$, *Ever enrolled* equals one if the student enrolled in type k in any year from 1 through 7. *Graduation* equals one if the student enrolled and graduated from type k in any year from 1 through 7; current enrollment without a recorded graduation does not count as graduation. *Still enrolled* equals one if the student is enrolled in type k in year 7 and has not graduated from type k by year 7. *Dropout* equals one if the student ever enrolled in type k by year 7, is not enrolled in any higher education institution in year 7, and has not graduated from any higher education institution by year

7. This corresponds to the case in which a student enrolls and subsequently drops out at any point during the seven years post high school graduation. Students who transfer from selective to non-selective institutions are not classified as dropouts.

We classify university majors as STEM or non-STEM according to the classification reported in Table A2. For the field-specific educational outcomes (e.g. Table A14), we define *Ever enrolled* in STEM as enrolling in a STEM program at any point during the first seven years after high school; *Graduation* in STEM as graduating from a STEM program within the first seven years after high school; *Dropout* from STEM as enrolling in a STEM program at any point during the first six years after high school but, by the seventh year, being neither enrolled in any degree program nor having graduated from any degree program; and *Still enrolled* in STEM as being enrolled in a STEM program in the seventh year after high school without having yet graduated in a STEM program. The outcomes for non-STEM programs are defined similarly.

For students who enroll in higher education, we also define program selectivity and relative academic position using the student's first observed institution–course enrollment (corresponding to an institution–major pair). *Selectivity* is the average standardized SIMCE score among students in the same institution–course pair and first-enrollment year, who did not enroll through preferential admissions. *Ability distance* is *Selectivity* minus the student's own standardized SIMCE score, measuring how far from the average peer the student is in terms of grade 10 standardized test scores. *Rank* is the student's within institution–course SIMCE rank, computed as the ordered SIMCE position divided by the number of students in the institution–course pair; the highest-ranked student has value 1, and ties receive the same rank value.

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".[41] 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 obtaining a PSU entrance exam score above an admission cutoff. Given that the vast majority of students in our sample are confident their PSU

[41]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.*

score *will* be high enough to obtain an admission (see the detailed analysis of pre-college beliefs in Tincani, Kosse, and Miglino (2026)), 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 the predicted probability of selective-college graduation, conditional on enrolling in a selective college. We predict each student's true probability of graduating from a selective college using a LASSO Probit regression estimated on the sample of students who have enrolled in selective college during the seven years post high school, using graduation from selective college within seven years of high school graduation 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 (0%, 25%, 50%, 75%, 100%). For robustness, we examine sensitivity to alternative numerical assignments in Figure A11.

B.2 Surrogate index

B.2.1 Methodology

To predict long-term effects on labor market outcomes we apply the surrogate index method described in Athey, Chetty, Imbens, and Kang (2025). We use data from five older cohorts of students to combine several shorter-run outcomes, such as educational performance and labor market outcomes up to seven years after high school, into a surrogate index, i.e., the predicted value of the longer-term outcome. Table A4 provides summary statistics for this sample of older cohorts.

To predict each long-term outcome and time horizon, such as average annual earnings over years 12–15, we estimate the following linear regression model using data on the older student cohorts:

$$Y_{is\tau} = \beta_{0\tau} + \beta_{1\tau} Z_i + \beta_{2\tau} X_{is} + u_{i\tau} \tag{11}$$

where $Y_{is\tau}$ is the labor-market outcome at time horizon τ 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 baseline student characteristics, school characteristics, cohort size, and labor-market conditions at the time of high school graduation. The parameters are indexed by τ because we estimate a separate regression for each time horizon.

Vector Z_i includes intermediate education and labor market outcomes measured over the first seven years post high school. The higher education outcomes include:

- high school GPA, an indicator for whether GPA is missing,
- annual indicators for enrollment in any type of higher education institution, one for each of the first seven years after high school,
- annual indicators for enrollment in a selective college (SUA), one for each of the first seven years after high school,
- graduation from any institution by year 7,
- graduation from a selective college (SUA) by year 7,
- for each annual enrollment indicator, interactions with the corresponding annual major-area indicator,
- for each annual enrollment indicator, interactions with gender,
- for each annual enrollment indicator, three-way interactions with the corresponding annual major-area indicator and gender,
- for each graduation-by-year-7 indicator, interactions with graduation major area,
- for each graduation-by-year-7 indicator, interactions with gender,
- for each graduation-by-year-7 indicator, three-way interactions with graduation major area and gender,
- annual indicators for dropout from any institution after prior enrollment, one for each of the first seven years after high school,
- annual indicators for dropout from a selective college (SUA) after prior enrollment, one for each of the first seven years after high school,
- for each annual dropout indicator, interactions with the corresponding annual indicator for the major from which the student dropped out,
- for each annual dropout indicator, interactions with gender,
- for each annual dropout indicator, three-way interactions with the corresponding annual indicator for the major from which the student dropped out and gender.

The labor market outcomes include:

- yearly months employed,
- a four-category annual education/employment status indicator, distinguishing individuals observed in higher education only, in private-sector employment only, in both, or in neither,
- interactions of yearly months employed with the occupational sector,
- interactions of yearly months employed with the annual education/employment status indicator,
- interactions of yearly months employed with gender,
- higher-order interactions among yearly months employed, the annual education/employment status indicator, the occupational sector, and gender,

– yearly earnings,
– interactions of yearly earnings with the occupational sector,
– interactions of yearly earnings with the annual education/employment status indicator,
– interactions of yearly earnings with gender,
– higher-order interactions among yearly earnings, the annual education/employment status indicator, the occupational sector, and gender.

Vector X_{is} includes student gender, age, an indicator for having never failed a grade, high school track (academic or vocational), the gender-specific youth unemployment rate in year 1 after high school graduation, school rurality, whether the school is in the Santiago metropolitan area, whether the school offers only academic tracks, and school-by-cohort size.

Armed with the estimated coefficients from the regressions in (11), 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.

B.2.2 Estimates

Tables B.I and B.II report selected coefficients from the surrogate-index models used to predict average annual earnings and monthly income in years 8–11 and 12–15. These coefficients are predictive rather than causal: they describe how the model maps observed baseline characteristics, education trajectories, and early labor-market outcomes into later labor-market outcomes. The tables show that baseline academic preparation and school context are important predictors. High school GPA is strongly positively associated with both predicted annual earnings and predicted monthly income at both horizons, while age is negatively associated with both outcomes. Prior labor-market outcomes are also highly predictive, especially earnings in year 7 after high school.

The education variables enter the model through enrollment, graduation, dropout, field-of-study indicators, and their interactions. The omitted field category is education, so the uninteracted enrollment, graduation, and dropout coefficients refer to students in the education-major reference category. Graduation from any higher education institution by year 7 in the reference category is positively associated with both predicted annual earnings and predicted monthly income. The displayed field interactions show that these associations differ substantially across majors, highlighting why the surrogate-index model allows education outcomes to enter flexibly rather than only through aggregate indicators. The model also allows gender differences to operate through field-specific and labor-market-history interactions, rather than only through the female indicator.

Overall, Tables B.I and B.II illustrate the main logic of the surrogate index: later labor-market outcomes are predicted using a rich combination of baseline preparation, higher-education progression, field of study, and early formal-sector labor-market attachment.

Table B.I: Surrogate-index coefficients for predicted annual earnings

	Years 8–11	Years 12–15
Female	333.061	-483.561
	(415.866)	(655.588)
Age	-26.261***	-49.663***
	(4.797)	(7.562)
Never failed	66.824***	121.262***
	(18.298)	(28.846)
Academic track	-37.993**	-32.623
	(17.550)	(27.667)
Gender-spec. unempl. Y1	10.307***	90.457***
	(1.716)	(2.704)
GPA	267.098***	476.777***
	(15.458)	(24.369)
Missing GPA	1133.107***	1698.656***
	(365.459)	(576.123)
Rural school	-39.907	-44.191
	(28.007)	(44.152)
Santiago school	145.378***	90.259***
	(14.349)	(22.620)
Academic-only school	-18.372	-5.797
	(16.122)	(25.416)
Cohort size	0.671***	1.646***
	(0.062)	(0.097)
Grad any by Y7	4608.642***	4100.255*
	(1345.808)	(2121.585)
Grad SUA by Y7	239.075	490.811
	(239.824)	(378.067)
Grad any Y7 × business/admin (non-STEM)	-3794.461***	-3861.260*
	(1334.407)	(2103.611)
Grad any Y7 × business/admin (non-STEM) × female	-36.751	369.397
	(284.246)	(448.096)
Grad any Y7 × engineering (STEM)	-1862.352	-1875.293
	(1331.635)	(2099.242)
Grad any Y7 × engineering (STEM) × female	-1263.994***	-2349.997***
	(398.222)	(627.772)
Grad any Y7 × arch./const. (non-STEM)	-2621.580*	-2133.141
	(1346.004)	(2121.893)
Grad any Y7 × arch./const. (non-STEM) × female	-1125.001**	-1210.002

Continued on next page

Continued from previous page

	Years 8–11	Years 12–15
	(552.805)	(871.464)
Grad any Y7 × health (STEM)	-4242.395***	-4389.678**
	(1337.700)	(2108.803)
Grad any Y7 × health (STEM) × female	-20.461	-24.732
	(293.739)	(463.061)
Enrolled any Y6	6782.653***	7451.058***
	(1590.796)	(2507.792)
Enrolled SUA Y6	35.405	267.763
	(247.975)	(390.918)
Dropout any Y7	1743.402	4093.822*
	(1372.218)	(2163.218)
Dropout any Y7 from business/admin (non-STEM)	-1919.827	-4327.843**
	(1347.084)	(2123.596)
Dropout any Y7 from business/admin (non-STEM) × female	-998.489**	-795.043
	(398.001)	(627.423)
Dropout any Y7 from engineering (STEM)	-2143.333	-4265.668**
	(1341.467)	(2114.740)
Dropout any Y7 from engineering (STEM) × female	278.198	-519.050
	(545.603)	(860.109)
Dropout any Y7 from arch./const. (non-STEM)	-1703.738	-4341.132**
	(1353.714)	(2134.048)
Dropout any Y7 from arch./const. (non-STEM) × female	-1388.846**	-1205.558
	(649.612)	(1024.074)
Dropout any Y7 from health (STEM)	-1852.436	-4170.683*
	(1365.272)	(2152.268)
Dropout any Y7 from health (STEM) × female	-423.288	-670.472
	(462.182)	(728.601)
Earnings Y5	0.056***	0.119***
	(0.018)	(0.028)
Earnings Y5 × female	0.045	-0.044
	(0.038)	(0.060)
Earnings Y7	0.740***	0.589***
	(0.012)	(0.019)
Earnings Y7 × female	-0.013	0.036
	(0.025)	(0.039)
Months worked Y5	14.415	-25.653
	(11.473)	(18.087)
Months worked Y5 × female	-24.229	26.104
	(20.869)	(32.898)
Months worked Y7	-87.937***	-56.105***
	(9.967)	(15.712)
Months worked Y7 × female	30.155*	-64.788**

Continued on next page

Continued from previous page

	Years 8–11	Years 12–15
	(17.497)	(27.583)
Earnings Y7 × transport	0.004	-0.016
	(0.021)	(0.033)
Earnings Y7 × transport × female	0.075*	0.028
	(0.044)	(0.070)
Earnings Y7 × professional/scientific	-0.040*	0.058*
	(0.021)	(0.034)
Earnings Y7 × professional/scientific × female	0.081**	-0.067
	(0.039)	(0.062)
Earnings Y7 × education sector	0.068*	0.048
	(0.036)	(0.056)
Earnings Y7 × education sector × female	0.051	0.042
	(0.045)	(0.070)

Notes: This table reports selected coefficients from OLS surrogate-index models used to predict average annual earnings, with missing Seguro earnings set to zero, in years 8–11 and 12–15 after high school. The table displays selected baseline controls, education outcomes, labor-market outcomes, and selected interactions with common major areas, common sectors, and female. The full models also include all remaining education outcomes from years 1–7, remaining major-area interactions, dropout-from-major interactions, gender interactions, labor-sector and education/labor-status indicators, and remaining interactions between labor outcomes, sector/status indicators, and gender. GPA is NEM. Gender-spec. unempl. Y1 is the gender-specific youth unemployment rate in the first post-high-school year. Santiago and Rural refer to the student's high school. SC denotes Seguro de Cesantia labor-market records. The base categories are only SC records for status, manufacturing for sector, and education major for enrollment, graduation, and dropout major indicators. Earnings and months-worked uninteracted coefficients refer to men in the only-SC status and manufacturing reference cell. Enrollment, graduation, and dropout uninteracted coefficients refer to men in the education-major reference category; these education coefficients are not interacted with status or sector. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Table B.II: Surrogate-index coefficients for predicted monthly income

	Years 8–11	Years 12–15
Female	85.462**	48.659
	(36.674)	(59.079)
Age	-2.383***	-5.283***
	(0.409)	(0.635)
Never failed	3.987**	8.977***
	(1.598)	(2.533)
Academic track	-1.695	-2.161
	(1.556)	(2.480)
Gender-spec. unempl. Y1	-0.031	0.664***

Continued on next page

Continued from previous page

	Years 8–11	Years 12–15
	(0.154)	(0.246)
GPA	35.780***	63.290***
	(1.378)	(2.197)
Missing GPA	280.793***	187.542
	(64.629)	(115.722)
Rural school	-5.424**	-3.237
	(2.460)	(3.912)
Santiago school	23.317***	13.979***
	(1.243)	(1.981)
Academic-only school	-2.842**	-2.240
	(1.408)	(2.244)
Cohort size	0.064***	0.147***
	(0.005)	(0.009)
Grad any by Y7	290.106***	691.990***
	(105.854)	(163.611)
Grad SUA by Y7	18.093	8.923
	(21.018)	(35.136)
Grad any Y7 × business/admin (non-STEM)	-173.512*	-503.285***
	(104.669)	(161.291)
Grad any Y7 × business/admin (non-STEM) × female	-18.219	13.517
	(24.626)	(39.766)
Grad any Y7 × engineering (STEM)	-138.884	-495.301***
	(104.389)	(160.781)
Grad any Y7 × engineering (STEM) × female	7.542	-85.529
	(34.041)	(54.548)
Grad any Y7 × arch./const. (non-STEM)	-75.792	-441.848***
	(105.828)	(163.132)
Grad any Y7 × arch./const. (non-STEM) × female	-87.103*	-7.627
	(48.782)	(79.296)
Grad any Y7 × health (STEM)	-146.641	-479.959***
	(105.629)	(162.981)
Grad any Y7 × health (STEM) × female	-16.401	-19.501
	(28.703)	(47.233)
Enrolled any Y6	396.102***	406.736**
	(126.467)	(195.890)
Enrolled SUA Y6	5.884	20.480
	(22.202)	(36.523)
Dropout any Y7	262.923**	549.840***
	(109.628)	(172.431)
Dropout any Y7 from business/admin (non-STEM)	-228.832**	-535.318***
	(107.367)	(168.641)
Dropout any Y7 from business/admin (non-STEM) × female	-52.272	-2.817

Continued on next page

Continued from previous page

	Years 8–11	Years 12–15
	(34.243)	(54.505)
Dropout any Y7 from engineering (STEM)	-259.459**	-526.727***
	(106.799)	(167.719)
Dropout any Y7 from engineering (STEM) × female	98.319**	42.704
	(48.580)	(74.713)
Dropout any Y7 from arch./const. (non-STEM)	-256.181**	-584.240***
	(107.914)	(169.498)
Dropout any Y7 from arch./const. (non-STEM) × female	-118.908**	-93.144
	(56.820)	(92.189)
Dropout any Y7 from health (STEM)	-242.097**	-521.951***
	(109.326)	(172.077)
Dropout any Y7 from health (STEM) × female	-46.032	-19.198
	(41.458)	(67.037)
Earnings Y5	0.006***	0.015***
	(0.001)	(0.002)
Earnings Y5 × female	0.002	-0.007
	(0.003)	(0.005)
Earnings Y7	0.061***	0.052***
	(0.001)	(0.001)
Earnings Y7 × female	-0.003	-0.000
	(0.002)	(0.003)
Months worked Y5	-1.212	-5.836***
	(0.914)	(1.441)
Months worked Y5 × female	-0.202	5.226*
	(1.729)	(2.763)
Months worked Y7	-24.099***	-19.616***
	(0.787)	(1.245)
Months worked Y7 × female	6.902***	1.525
	(1.395)	(2.264)
Earnings Y7 × transport	-0.002	-0.005*
	(0.002)	(0.003)
Earnings Y7 × transport × female	0.007**	0.007
	(0.003)	(0.006)
Earnings Y7 × professional/scientific	-0.000	0.004*
	(0.002)	(0.003)
Earnings Y7 × professional/scientific × female	0.006**	0.002
	(0.003)	(0.005)
Earnings Y7 × education sector	0.005*	-0.007
	(0.003)	(0.005)
Earnings Y7 × education sector × female	0.001	0.008
	(0.004)	(0.006)

Continued on next page

Continued from previous page

	Years 8–11	Years 12–15

Notes: This table reports selected coefficients from OLS surrogate-index models used to predict average monthly income in years 8–11 and 12–15 after high school. The table displays selected baseline controls, education outcomes, labor-market outcomes, and selected interactions with common major areas, common sectors, and female. The full models also include all remaining education outcomes from years 1–7, remaining major-area interactions, dropout-from-major interactions, gender interactions, labor-sector and education/labor-status indicators, and remaining interactions between labor outcomes, sector/status indicators, and gender. GPA is NEM. Gender-spec. unempl. Y1 is the gender-specific youth unemployment rate in the first post-high-school year. Santiago and Rural refer to the student's high school. SC denotes Seguro de Cesantia labor-market records. The base categories are only SC records for status, manufacturing for sector, and education major for enrollment, graduation, and dropout major indicators. Earnings and months-worked uninteracted coefficients refer to men in the only-SC status and manufacturing reference cell. Enrollment, graduation, and dropout uninteracted coefficients refer to men in the education-major reference category; these education coefficients are not interacted with status or sector. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

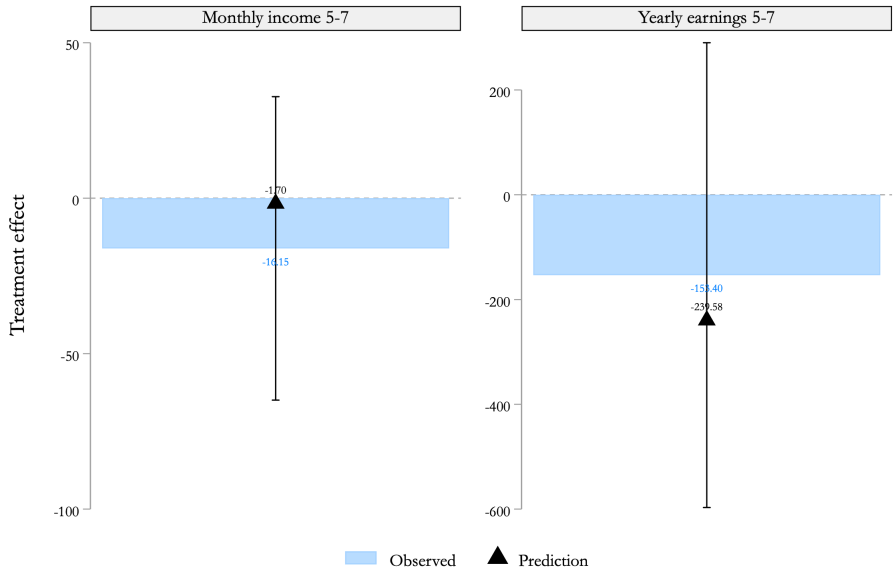
B.2.3 Supporting evidence

This exercise assesses whether this prediction procedure reproduces observed treatment effects in periods for which actual labor-market outcomes are already available.

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)[42]; 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. (2025), we provide supporting evidence in favour of the third assumption by comparing the impact estimates on annual earnings and monthly income averaged over years five to seven post high school, which are observed, to those obtained using the surrogate index for year five to seven earnings and monthly income calculated using data from the first four years after high school. The results of this exercise are reported in Figures B.1 and B.2, which show a close alignment between actual and predicted effects for both RDD and RCT samples, and with predicted effects always lying within the 95% confidence interval for the observed effects.

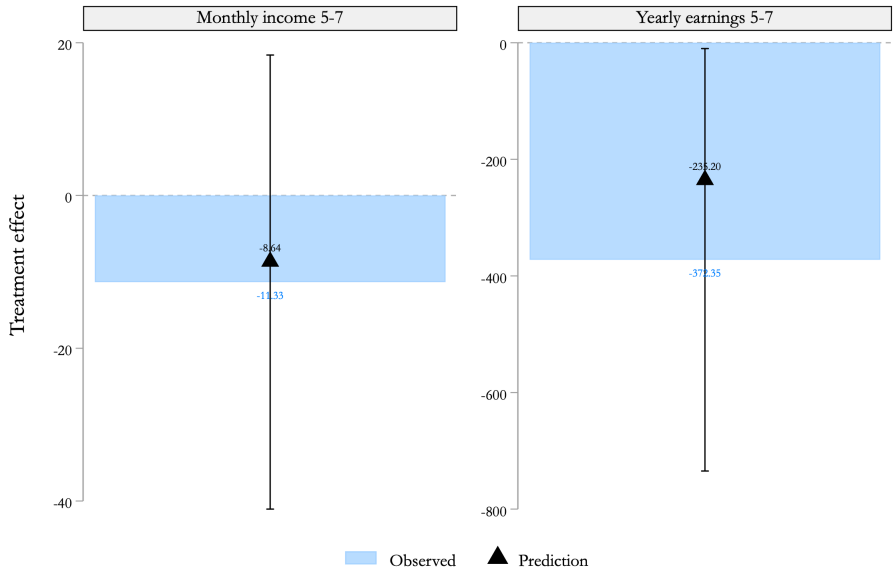
[42]This assumption can be applied to the RDD setting under a local randomization assumption: the treatment is 'as good as random' in a small window around the cutoff.

Figure B.1: Supporting evidence: predictions on labor-market outcomes, RD sample



Notes: This figure shows the estimate of parameter δ in equation (2) in the RDD sample of students around the within-school top 15% cutoff. *Observed* are the point estimates when the dependent variable is the observed labor market outcome averaged over years five to seven post high school. *Predictions* are the point estimates when the dependent variable is the predicted labor market variable averaged over the same period, where the prediction is performed using observed data up to year 4 after high school. The outcome variable in the left panel is monthly income and in the right panel annual earnings. The bars represent 95% confidence intervals for the treatment effect on the observed outcomes, calculated from standard errors clustered at the school level. The controls are the standard set: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, type of high school track (academic or vocational), and school fixed effects.

Figure B.2: Supporting evidence: labor-market outcomes, experimental sample (baseline top 15%)



Notes: This figure shows the estimate of parameter β in equation (1) in the experimental sample of students who belong to the top 15% of their baseline GPA in their school. *Observed* are the point estimates when the dependent variable is the observed labor market outcome averaged over years five to seven post high school. *Predictions* are the point estimates when the dependent variable is the predicted labor market variable averaged over the same period, where the prediction is performed using observed data up to year 4 after high school. The outcome variable in the left panel is monthly income and in the right panel annual earnings. The bars represent 95% confidence intervals for the treatment effect on the observed outcomes, calculated from standard errors clustered at the school level. The controls are the standard set: gender, age, socioeconomic status, baseline standardized test scores, whether they ever repeated a grade, type of high school track (academic or vocational), and school-region fixed effects.

B.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. To obtain bounds on the RD estimates, we adapt the Lee bounds procedure Lee (2009) 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. Let $\hat{\delta}_{\text{Enr}}$ denote the RD estimate of δ in Equation (2), using an indicator for higher-education enrollment in the first year after high school as the dependent variable. A PACE admission (defined as $A_i = 1$) leads to a $\hat{\delta}_{\text{Enr}}$ percentage point increase in the enrollment rate and $\tau = \frac{\hat{\delta}_{\text{Enr}}}{E(Enrollment_{is}|r_{is}^{+}=0,A_i=1)}$ is the ratio of the treatment effect on enrollment and the enrollment rate of the treated group at the cutoff.

Then, an upper bound of the RD estimate for outcome Y (e.g. selectivity) 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_{is} 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. After trimming, we estimate the fuzzy RD specification for outcome Y on the restricted sample. Under the continuity, monotonicity, and smoothness assumptions of Dong (2019), the resulting estimates provide upper and lower bounds on the local average treatment effect of a PACE admission on outcome Y.

B.4 RDD estimations away from the cutoff

B.4.1 Method and assumptions

The RDD estimates in Section 5.3 identify the effect of preferential admission for students close to the PACE admission-eligibility cutoff. In Section 6.1, we use the method of Angrist and Rokkanen (2015) to estimate the effects of eligibility for students who are farther from the admission margin. In our setting, the treatment is eligibility for a PACE admission offer.

Let $r_{is} = p_{is} - c_s$ denote the running variable for student i in school s, centered at the school-specific PACE cutoff. Students with $r_{is} \geq 0$ are eligible for preferential admission, while students with $r_{is} < 0$ are not. Because cutoffs are school-specific, we stack students across schools after normalizing the running variable relative to each school cutoff. The extrapolation is conducted among students for whom the running variable is observed, which requires taking the entrance exam.

The method relies on two assumptions. First, potential outcomes with and without preferential-admission eligibility must be mean independent of the running variable conditional on a vector of covariates X, the conditional independence assumption. Second, conditional on X, students must have positive probability of being observed on both sides of the eligibility cutoff, the common-support assumption. Intuitively, the method can extrapolate away from the cutoff only for students whose covariates are represented among both eligible and ineligible students.

Figure B.4 assesses the plausibility of the conditional independence assumption. It plots the relationship between each outcome and the running variable separately on either side of the cutoff, before and after conditioning on grade-12 GPA for graduation, dropout and monthly income and grade-11 GPA, average grade-10 and grade-9 GPA for average annual earnings. The conditional relationships are approximately flat for graduation from a selective college, dropout from any institution, average annual earnings over years 1–7, and monthly income in year 7. This provides evidence consistent with the conditional

independence assumption for the baseline conditioning set.[43] Figure B.5 assesses common support by plotting the distribution of the estimated propensity score $P(A_i = 1 \mid X_i)$, where $A_i = 1$ denotes being above the top 15% cutoff. The distributions for students above and below the cutoff show substantial overlap, consistent with the common-support requirement.

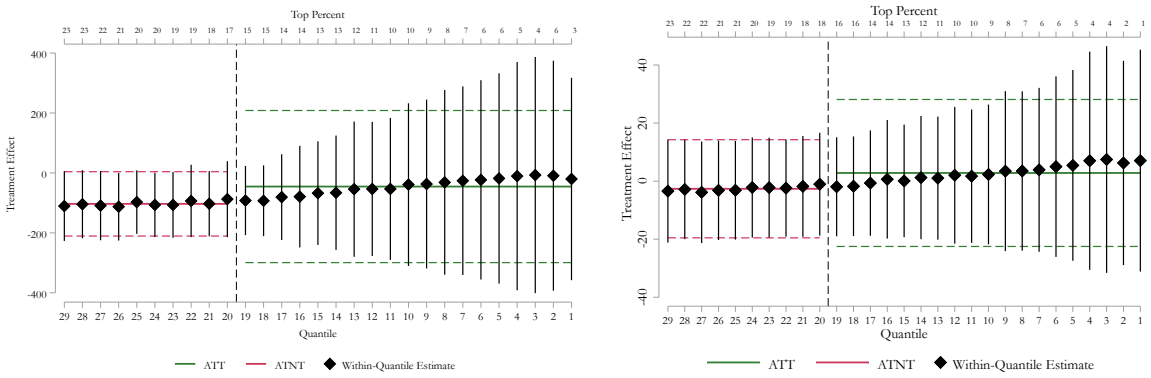
To construct Figure 8 in the main text, we restrict the sample to students approximately in the top 25% of the within-school high school GPA distribution and for whom the PRN running variable is observed. We divide the sample to the right of the top 15% cutoff into 19 equally sized groups and the sample to the left into 10 equally sized groups, ordered by distance from this cutoff. For each group, we estimate the effect of preferential-admission eligibility using the Angrist and Rokkanen (2015) extrapolation procedure. Figure 8 reports the resulting estimates for graduation from a selective college and dropout from any higher education institution, the two outcomes that summarize the main access–mismatch tradeoff.

The mapping from extrapolation groups to policy-relevant GPA ranks requires one additional step. The PRN running variable is available to researchers only for students who take the entrance exam, while the policy-relevant ranking is defined within the full high school cohort. We therefore map each extrapolation group among entrance-exam takers into its corresponding percentile of the school-wide high school GPA distribution. To do so, we calculate the average within-school GPA rank, considering everyone in the school, of students belonging to each extrapolation group. In Figure 8, the first group for which the estimated effect on dropout from any higher education institution is positive and statistically significant at the 5% level corresponds approximately to the 94^{th} percentile of the school-wide GPA distribution.

This mapping implies that, among students who take the entrance exam, dropout risk becomes statistically detectable once preferential-admission eligibility is extended to students ranked lower than roughly the top 6% of their high school cohort. More inclusive rules, such as top-7% or top-8% rules, would extend eligibility farther down the achievement distribution, into the range where the extrapolated estimates detect positive dropout effects. This interpretation requires two additional assumptions: first, that changing the eligibility cutoff would not alter entrance-exam taking as a function of within-school GPA rank; and second, that non-exam takers at a given rank would have dropout outcomes similar to exam takers at that rank.

[43] Grade-12 GPA does not coincide with the running variable. The PACE running variable is based on GPA rank over the four years of high school and includes additional adjustments; see Tincani, Kosse, and Miglino (2026) for the PRN formula.

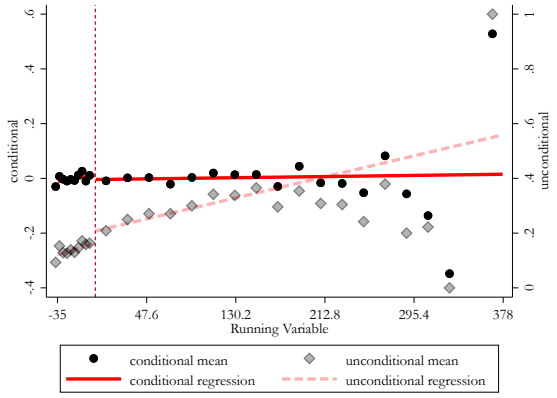
Figure B.3: RDD away from the cutoff



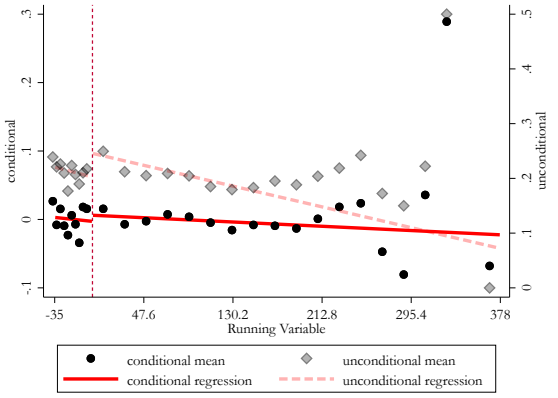
(a) Average annual earnings years 1-7

(b) Average monthly income year 7

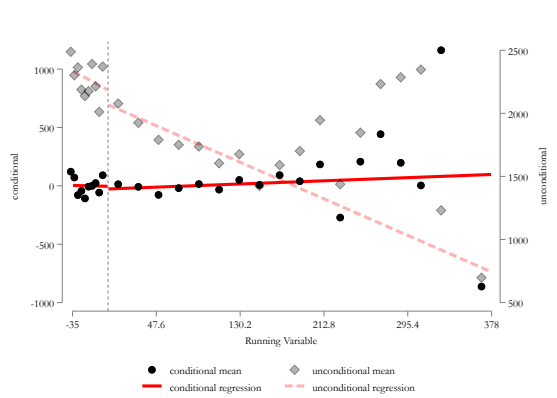
Notes: This figure reports estimates of treatment effects away from the cutoff, following Angrist and Rokkanen (2015). The conditioning vector consists of grade-11 GPA, average grade-10 and grade-9 GPA for average annual earnings and of grade-12 GPA for average monthly income. ATT refers to the average treatment effect on eligible students, and ATNT to the average treatment effect on ineligible students. Vertical bars correspond to 95% confidence intervals, obtained from standard errors clustered at school level computed using a nonparametric bootstrap with 100 replications. The lower horizontal axis indexes quantiles of the running variable distribution. The upper horizontal axis reports, for each group, the average school-wide GPA ranking, calculated among all students in the school and based on GPA in all four high school years. These rankings are rounded to the nearest integer, so two distinct groups may display the same integer cutoff when their exact values fall in the same rounding interval.



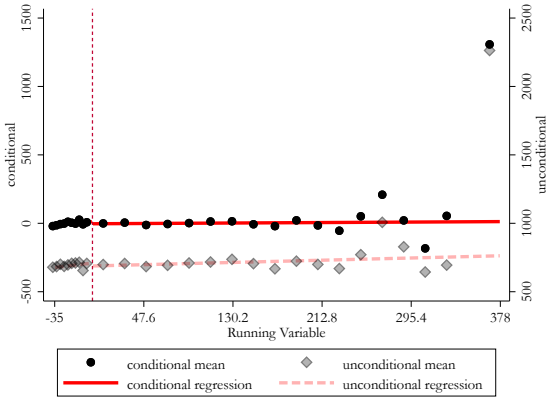
(a) Graduation from a selective college



(b) Dropout from any institution

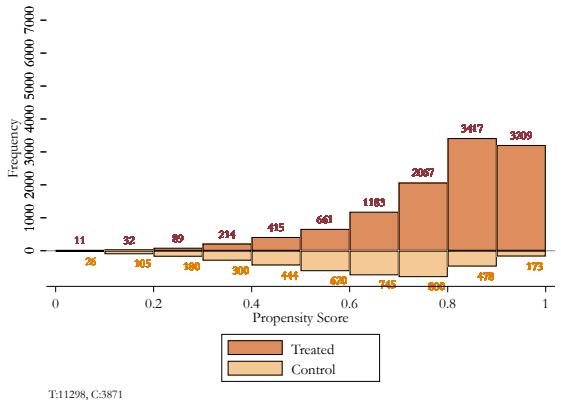


(c) Average annual earnings years 1-7

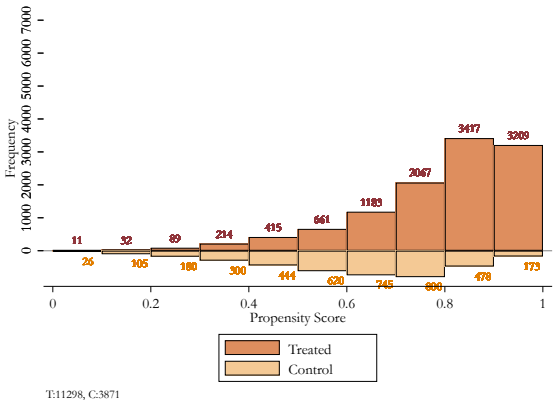


(d) Monthly income year 7

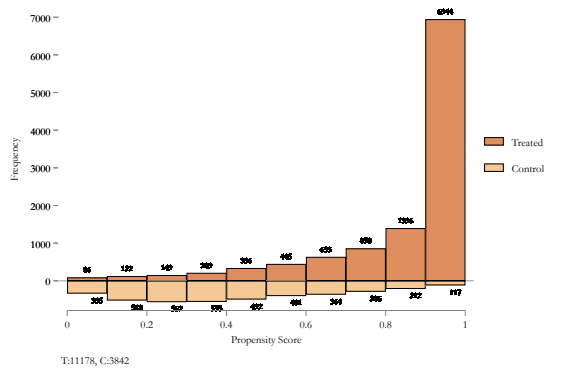
Figure B.4: Evidence on the Conditional Independence Assumption. Each panel shows, for the corresponding outcome, the relationship between the outcome and the running variable separately on either side of the cutoff, before and after conditioning on grade-12 GPA for graduation, dropout and monthly income and grade-11 GPA, average grade-10 and grade-9 GPA for average annual earnings. Under the CIA, the conditional relationship should be flat within each side of the cutoff. P-values for the slopes (left and right of the cutoff): (a) L: 0.19 R: 0.39 (b) L: 0.79 R: 0.15 (c) L: 0.96 R: 0.19 (d) L: 0.34 R: 0.53.



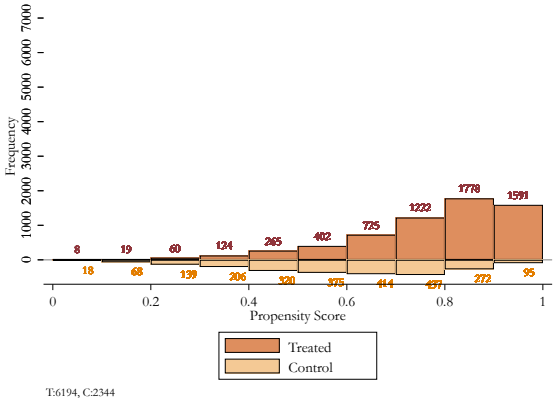
(a) Graduation from a selective college



(b) Dropout from any institution



(c) Average annual earnings years 1-7



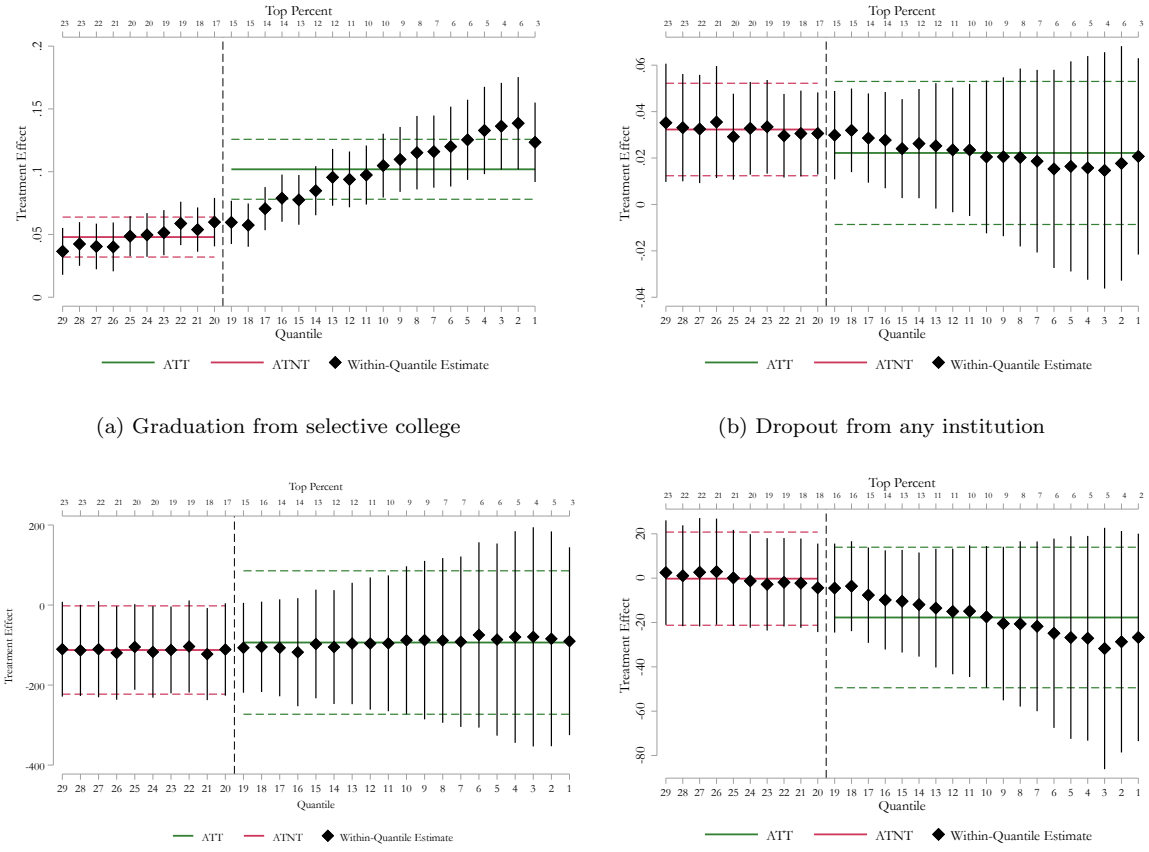
(d) Monthly income year 7

Figure B.5: Assessment of Common Support. Each panel plots the distribution of the propensity score $P(D = 1 \mid X)$, estimated through a Logit model, where $D = 1$ denotes being above the top 15% cutoff, separately for students above and below the cutoff. Overlap in these distributions provides evidence of common support for the CIA-based extrapolation.

B.4.2 Robustness

Figure B.6 reports robustness checks using an alternative conditioning vector that includes grade-10 SIMCE, grade-10 GPA, and grade-11 GPA. The results preserve the main qualitative pattern: preferential-admission eligibility continues to generate gains in selective-college graduation away from the cutoff, while positive effects on dropout emerge as eligibility expands farther down the academic ranking. The precise group at which dropout becomes statistically detectable differs across conditioning sets, so we interpret the top-6% threshold as an approximate policy-relevant benchmark rather than an exact cutoff.

Figure B.6: RDD Away From the Cutoff: Robustness to Alternative Conditioning Set



(a) Graduation from selective college

(b) Dropout from any institution

(c) Average earnings years 1-7

(d) Monthly income year 7

Notes: This figure reports estimates of treatment effects away from the cutoff, following Angrist and Rokkanen (2015). The conditioning vector consists of grade-10 SIMCE test score, grade-10 GPA, and grade-11 GPA for graduation, dropout and average annual earnings and grade-10 SIMCE test score, grade-12 GPA, and grade-9 GPA for monthly earnings. ATT refers to the average treatment effect on eligible students, and ATNT to the average treatment effect on ineligible students. Vertical bars correspond to 95% confidence intervals, obtained from standard errors clustered at school level computed using a nonparametric bootstrap with 100 replications. The lower horizontal axis indexes quantiles of the running variable distribution. The upper horizontal axis reports, for each group, the average school-wide GPA ranking, calculated among all students in the school and based on GPA in all four high school years. These rankings are rounded to the nearest integer, so two distinct groups may display the same integer cutoff when their exact values fall in the same rounding interval.

B.5 Empirical Decomposition Result: Additional Details

Let $C_i = 1$ denote the common-support region. The Angrist and Rokkanen (2015) estimator identifies $ATT_A^{CS} = E[Y_{11} - Y_{10} \mid A_i = 1, S_i = 1, C_i = 1]$. Equation (10) uses the full eligible-population ATT_A. We interpret ATT_A^{CS} as ATT_A under the additional overlap external-validity condition that the mean effect of preferential-admission eligibility for eligible students outside common support equals the mean effect for eligible students inside common support, $E[Y_{11} - Y_{10} \mid A_i = 1, S_i = 1, C_i = 0] = E[Y_{11} - Y_{10} \mid A_i = 1, S_i = 1, C_i = 1]$, or differs by too little to affect the residual term. The share of observations

outside common support in the sample underlying Table 5 is 18.5% for selective-college graduation and 9.1% for dropout from any institution.

C Theoretical Appendix

Proof of Lemma 1. Take the individual-level effect of eligibility for preferential admission if in PACE school, $Y_{11} - Y_{10}$, and subtract and add Y_{00}:

$$\begin{aligned} \Delta_a &= Y_{11} - Y_{10} \\ &= (Y_{11} - Y_{00}) - (Y_{10} - Y_{00}) \end{aligned} \tag{12}$$

Rearranging:

$$\underbrace{Y_{11} - Y_{00}}_{\substack{\text{effect of PACE school} \\ \text{with eligibility}}} = \underbrace{(Y_{11} - Y_{10})}_{\substack{\text{effect of eligibility} \\ \text{cond on PACE school}}} + \underbrace{(Y_{10} - Y_{00})}_{\substack{\text{effect of PACE school} \\ \text{without eligibility}}}$$

■

Proof of Theorem 1. Using the definition of Y_{1*} and rearranging,

$$Y_{1*} - Y_{00} = a(Y_{11} - Y_{10}) + (Y_{10} - Y_{00}).$$

Taking conditional expectations:

$$CATE_S = E[Y_{1*} - Y_{00}|X] = Pr(a = 1|X)E[Y_{11} - Y_{10}|X, a = 1] + E[Y_{10} - Y_{00}|X]. \tag{13}$$

By Assumption 1,

$$Pr(a = 1|X) = Pr(a = 1|X, S = 1) = \pi_X. \tag{14}$$

By Assumption 2,

$$E[Y_{11} - Y_{10}|X, a = 1] = E[Y_{11} - Y_{10}|X, a = 1, S = 1]. \tag{15}$$

The admission eligibility status in PACE schools is observed, so that in large samples the term on the right-hand side is known and equal to $E[Y_{11} - Y_{10}|X, A = 1, S = 1]$. Therefore, it follows that:

$$\underbrace{E[Y_{1*} - Y_{00}|X]}_{\text{CATE}_S} = \underbrace{\pi_X}_{\text{Cond. fraction eligible}} \underbrace{E[Y_{11} - Y_{10}|X, A = 1, S = 1]}_{\text{CATT}_A} + \underbrace{E[Y_{10} - Y_{00}|X]}_{\text{Conditional average effect of PACE school without eligibility}}. \tag{16}$$

■

Proof of Theorem 2. Integrating out X from the left- and right-hand sides of equation (16), we obtain:

$$\underbrace{\int E[Y_{1*} - Y_{00}|X]f(X)dX}_{\text{ATE}_S} = \int \pi_X E[Y_{11} - Y_{10}|X, A = 1, S = 1]f(X)dX + \int E[Y_{10} - Y_{00}|X]f(X)dX \tag{17}$$

By Assumption 3, we can rewrite the first term on the right-hand side as:

$$\int \pi_X E[Y_{11} - Y_{10}|X, A = 1, S = 1]f(X|S = 1)dX \tag{18}$$

Recall that $\pi_X = Pr(A = 1|X, S = 1)$. By the definition of joint conditional density, $Pr(A = 1|X, S = 1)f(X|S = 1) = f(X, A = 1|S = 1)$. Dividing both sides of this equation by $P(A = 1|S = 1)$:

$$\frac{Pr(A = 1|X, S = 1)f(X|S = 1)}{P(A = 1|S = 1)} = \frac{f(X, A = 1|S = 1)}{P(A = 1|S = 1)} \tag{19}$$

and by the definition of conditional probability, the right-hand side of equation (19) is equal to $f(X|S = 1, A = 1)$. Therefore, denoting $P(A = 1|S = 1)$ by π:

$$f(X|A = 1, S = 1) = \frac{\pi_X f(X|S = 1)}{\pi}$$

and rearranging:

$$\pi_X = \pi \cdot \frac{f(X|A = 1, S = 1)}{f(X|S = 1)}. \tag{20}$$

Using this expression for π_X, the first term on the right-hand side of equation (17) becomes:

$$\pi \cdot \int E[Y_{11} - Y_{10}|X, A = 1, S = 1] \frac{f(X|A = 1, S = 1)}{f(X|S = 1)} f(X|S = 1)dX =$$
$$= \pi \cdot \int E[Y_{11} - Y_{10}|X, A = 1, S = 1] f(X|A = 1, S = 1)dX \tag{21}$$

so that equation (17) can be rewritten as:

$$\underbrace{\int E[Y_{1*} - Y_{00}|X] f(X)dX}_{\text{ATE}_S} = \underbrace{\pi}_{\text{Fraction eligible}} \underbrace{\int E[Y_{11} - Y_{10}|X, A = 1, S = 1] f(X|A = 1, S = 1)dX}_{\text{ATT}_A} +$$
$$+ \underbrace{\int E[Y_{10} - Y_{00}|X] f(X)dX}_{\text{ATE of PACE school without eligibility}}$$

■

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.[44]

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 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.

[44]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.