Abhijit Banerjee
Esther Duflo
Francisco Gallego

Removing barriers to higher education in Chile

Evaluation of peer effects and scholarships for test preparation

March 2016

Impact Evaluation Report 36

Education

International Initiative for Impact Evaluation

Removing barriers to higher education in Chile

Evaluation of peer effects and scholarships for test preparation

Abhijit Banerjee
Department of Economics, Massachusetts Institute of Technology

Esther Duflo
Department of Economics, Massachusetts Institute of Technology

Francisco Gallego
Institute of Economics, Pontifical Catholic University of Chile

3ie Impact Evaluation Report 36

March 2016

International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making NGO promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, how, why and at what cost. We believe that better and policy-relevant evidence will make development more effective and improve people's lives.

3ie Impact Evaluations

3ie-supported impact evaluations assess the difference a development intervention has made to social and economic outcomes. 3ie is committed to funding rigorous evaluations that include a theory-based design, use the most appropriate mix of methods to capture outcomes and are useful in complex development contexts.

About this report

3ie accepted the final version of this report, *Removing higher education barriers of entry: test training and classroom peer effects in Chile,* as partial fulfilment of requirements under grant OW2.167 issued under Open Window 2. A copy-edited and formatted version of this report is now published as *Removing barriers to higher education in Chile: evaluation of peer effects and scholarships for test preparation, 3ie Impact Evaluation Report 36*. All of the content is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or 3ie Board of Commissioners. Any errors and omissions are also the sole responsibility of the authors. All affiliations of the authors listed in the title page are those that were in effect at the time the report was accepted. Any comments or queries should be directed to the corresponding author, Francisco Gallego at fgallego@uc.cl

Funding for this impact evaluation was provided by 3ie's donors, which include UKaid, the Bill & Melinda Gates Foundation, Hewlett Foundation and 12 other 3ie members that provide institutional support. A complete listing is provided on the 3ie website at http://www.3ieimpact.org/about-us/3ie-members/

Suggested citation: Banerjee, A, Duflo E, Gallego, F, 2016. *Removing barriers to higher education in Chile: evaluation of peer effects and scholarships for test preparation, 3ie Impact Evaluation Report 36.* New Delhi: International Initiative for Impact Evaluation (3ie)

3ie Impact Evaluation Report Series executive editors: Jyotsna Puri and Beryl Leach
Managing editor: Deepthy Menon and Omita Goyal
Assistant managing editors: Radhika Menon and Marcus Olapade
Production manager: Pradeep Singh
Copy editor: Aruna Ramachandran
Proofreader: James Middleton
Cover design: John F McGill
Printer: Via Interactive
Cover photo: Charlotte Kesl/The World Bank

Acknowledgements

Randomized evaluations, such as this, require the contributions of a large number of people. While it would be impossible to recognize everyone who made a contribution to this project, we would like to thank Michael Eddy, Elizabeth Weber, Noam Angrist, José I. Cuesta, Sebastian Otero, Josefa Aguirre and Alejandro Saenz for superb research assistance, Ryan Cooper for help with the implementation of the evaluation, and several members of Preuniversitario UC (our partner organization) for their support, and for providing data and qualitative information on the project. We are also grateful to the International Initiative for Impact Evaluation, Inc. (3ie), which provided funding support through the Global Development Network. The views expressed in this article are not necessarily those of 3ie or its members.

Abstract

Access to the most profitable tracks in higher education presents a steep socioeconomic gradient in most developed and developing countries. One potential explanation for this is performance on standardized tests, an integral component of the college admissions process in many countries, in which wealthy students consistently outperform poor students, presumably because of their better educational background.

This paper reports the results of a randomized experiment designed to evaluate the effects of offering scholarships on test outcomes and entry to higher education in Chile. These scholarships were offered for high-quality test preparation to high-achieving students from low socioeconomic backgrounds. The experiment took place in two different years, with big student protests affecting the second year.

We find that the scholarships significantly increased test preparation in high-quality institutions, especially among students from low-performing high schools. However, in terms of outcomes, while we find small and non-significant average impacts on test scores and higher education entry in the first cohort of students, we find significant impact for treated students in the second cohort, especially among students from low-performing schools.

Our interpretation is that over this period, in which students lost several school days, the scholarships provided a good option that substituted for preparation in school. This group of students also saw significant increases in the quality of higher education they accessed, entering more selective tracks and with significantly higher expected labor market outcomes.

In addition, we find evidence to suggest that the scholarships affected the opinions and behaviors of students related to the education policy discussion. Finally, we overlap a peer effects experiment in which scholarship students are allocated either to mixed-ability or tracking-by-ability classes. We do not find significant effects of this intervention on (blindly allocated) teachers and students outcomes.

Contents

List of figures and tables

Appendix tables

Abbreviations and acronyms

CFT	Centros de Formación Técnica (Centers of Technical Education)
CRUCH	Consejo de Rectores de la Universidades Chilenas (Chilean Universities Presidents Council)
DEMRE	Departmento de Evaluación, Medición y Registro Educational (Department of Evaluation, Measurement, and Educational Enrollment)
GPA	grade point average
IP	Institutos Profesionales (Professional Institutes)
IV	instrumental variable
ITT	intention-to-treat
LATE	local average treatment effect
MINEDUC	Ministry of Education
OLS	ordinary least squares
p.p.	percentage points
Preu UC	Preuniversitario UC
RCT	randomized controlled trial
RDD	regression discontinuity design
SIMCE	Sistema de Medición de Calidad de la Educación (Educational Quality Measurement System)

1 Introduction

Access to higher education has broad implications for development. A significant literature on returns to higher education shows large effects on earnings and productivity (Wachtel 1975; Katz and Murphy 1992; Blundell *et al.* 2000; Barrow and Rouse 2005; Cellini and Chaudhary 2012; Lang and Weinstein 2012). However, we consistently observe large barriers to higher education access. Indeed, poor students are significantly less likely to attend higher education institutions, even when the returns to education are high for disadvantaged students (Blossfeld and Shavit 1993; Ellwood and Kane 2000; Card 2001; Hastings, Neilson and Zimmerman 2013).

A relevant component of these barriers of access to higher education includes performance on standardized tests, an integral component of the college admissions process in many countries (Table 1 lists several countries using university entrance exams), in which wealthy students consistently outperform poor students, presumably because of their better educational background (Zwick 2012).

In theory, standardized tests level the playing field and increase access to higher education by making socioeconomic status and other individual characteristics less relevant (especially in contrast to more discretionary systems in which other elements such as essays and letters of recommendation play a role). To this end, standardized tests provide an avenue for low-income students to show their merit on a nationally comparable scale.

Moreover, standardized tests are often less susceptible to manipulation by students with more influence, networks, or money, and are thus promoted as a transparent and efficient means by which to allocate access to higher education. However, if wealthy students have differentiated access to costly, high-quality test preparation by being less financially constrained, or because test scores require skills acquired in primary and secondary school, entry into higher education may remain unequal.

Thus, differentiated access to test preparation takes place, either because some students can attend institutions set to that purpose, or because they have access to better-quality schools. In such a scenario, test preparation opportunities become systematically biased towards wealthier students, with test training reversing the potential benefits of standardized testing and harming the process of human capital accumulation, making it socially less efficient. Such a scenario would imply the existence of large barriers to access to higher education for poorer students.

In light of these tradeoffs, two questions emerge. First, are financial constraints a relevant barrier for test preparation? And, second, is test preparation effective in increasing standardized test performance and, if so, in increasing access to higher education? These are the two main questions we intend to answer in this paper, using evidence from Chile. Chile serves as a particularly relevant case study to evaluate these hypotheses, because it exhibits the previously mentioned characteristics of a context where access to higher education is assigned through a standardized test (an SAT-like test, the *Prueba de Selección Universitaria* [University Selection Test]

[PSU]), and also shows it to be strongly correlated with students' socioeconomic backgrounds (Reyes, Rodríguez and Urzúa 2013).

Indeed, in 2011, only 22 per cent of 18–25 year-olds from the lowest decile in Chile were enrolled in higher education, whereas 63 per cent from the highest decile were enrolled (CASEN 2011). Moreover, accessing the more profitable programs of higher education in Chile seems to produce sizable returns for students with different socioeconomic backgrounds (Hastings, Neilson and Zimmerman 2013; Reyes, Rodríguez and Urzúa 2013).

Finally, the case of Chile is also interesting because of a private market of test-preparing institutions called *preuniversitarios* that offer a wide range of alternatives, with heterogeneous qualities, forms and prices, and without much public information on these attributes. We investigate the impacts of test preparation on access to higher education using experimental evidence from a randomized controlled trial (RCT), which randomly offered scholarships for attending a high-quality test-training institution to low- and middle-income secondary school students (attending subsidized schools) with above average academic performance and who applied for the program.

The program offered a total of 1,284 partial scholarships – that covered about 65 per cent of the *preuniversitario* cost – for 12th graders of two different cohorts to attend Preuniversitario UC (Preu UC), a high-quality test preparation institution.[1] The cost of a typical package of test preparation in the top *preuniversitarios* is above US$1,500 per year (as a reference, the annual per capita income of a family located in the fifth percentile was about US$3,200 in 2011). The experiment took place in two different years, 2010 and 2011. Many student protests took place in the second year, which resulted in a significant loss of school days.

Results show that the scholarship offers in effect lowered financial barriers for test preparation and increased access to high-quality test preparation. Being offered a scholarship to attend Preu UC increases the likelihood of attending and completing Preu UC by 65.6 percentage points (p.p. hereafter), and increases the likelihood of attending a top *preuniversitario* by 37 p.p. (from a mean of about 50 per cent in the control group).

However, we observe significant heterogeneity among students from different cohorts, and among students attending high- and low-quality schools. While takeup in top preparation institutions increased by 45.2 p.p. for students of low-performing schools who were offered scholarships, the same variable increased by just 27.9 p.p. for students from high-performing schools. This is because students from high-performing schools had better access to high-quality test preparation; whereas in the control group 61 per cent of students from high-performing schools attended a top *preuniversitario*, 40 per cent of students from low-performing schools attended one.

[1] Three *preuniversitarios* are widely recognized as being of the highest quality among the available supply of massive *preuniversitarios*. These are CEPECH, Pedro de Valdivia, and Preuniversitario UC. We henceforth refer to these three institutions throughout this paper simply as the top *preuniversitarios*.

In relation to the role of private providers in this context; our evidence suggests that the higher proportion of students from high-quality schools attending high-quality preparation institutions is not a consequence of their being less financially constrained, but rather, that they have greater access to full or partial scholarships. (Although we do not have data to test how many students in our experiment had access to scholarships, we provide evidence about students' expenditure on *preuniversitarios*.) Our main hypothesis is that these institutions practice statistical discrimination and offer scholarships to high-achieving students from high-performing schools to improve the average performance of their students and use this as a marketing tool.

The effects for students from the two cohorts studied also differ. The increase in takeup as a result of our experiment is significantly lower for the second cohort, especially among students from low-performing schools. While the experiment increased the takeup by 51 p.p. for students from low-performing schools in the 2010 cohort, it increased the takeup by only 34 p.p. for students from low-performing schools in 2011. Furthermore, in the 2011 cohort the increase in takeup is not statistically different between students from low- and high-performing schools. Although we do not have systematic quantitative information about the causes for this decrease in the second cohort, we present evidence collected in our follow-up survey that suggests that the top test preparation institutions are increasingly entering new market segments with new scholarships and price discounts for high-ability students from low-performing schools.

Next we present intention-to-treat (ITT) estimates of our experiment (i.e. the impact of being offered a scholarship for a specific top *preuniversitario*) on educational outcomes. This experiment takes advantage of the random assignment of the experiment and compares directly the outcomes of those offered a scholarship with those not offered one.

Our first finding is that under regular circumstances (i.e. a year without student protests), the offer of the scholarship had no significant impact on students' results. Interestingly, receiving a scholarship led students to substitute school work with *preuniversitario* work, and overall there were no gains in students' performance. Our second finding is that under extraordinary circumstances (i.e. a year with student protests), in which students suffered the consequences of lost school days due to protests, the offer had sizable impacts. In fact, within our 2011 cohort, students increased their language and math test scores (the two mandatory sections of the PSU).

Notably, these impacts are concentrated in students from low-quality high schools, with students being offered a scholarship increasing their test scores in 0.13 standard deviations (hereafter σ) and 0.17σ. A similar pattern of results is found in terms of impacts of the scholarship offers on access to higher education: being offered the scholarship increased access by 15 p.p. (equivalent to 0.32σ). We also find significant impacts in terms of the quality of the higher education track students from low-performing schools in the second cohort choose: increasing their probability of entering a university track (and proportional decreases in the probability of entering a vocational track), attending more selective institutions (measured by the average PSU scores of the students entering the program), and attending programs with higher expected wages.

Given that the scholarship program resulted in a significant takeup increase, we also use this as an instrument for test preparation at a top *preuniversitario* and estimate the local average treatment effect of high-quality test preparation on standardized test performance, access to higher education and other related outcomes. The results mimic our ITT estimates but with stronger effects: within our 2011 cohort, the instrumental variable (IV) estimates imply that students attending a top *preuniversitario* increased language and math test scores by 0.49σ and 0.26σ, and increased higher education attendance by 42 p.p. (equivalent to in 0.94σ).

Our IV estimates capture a local average treatment effect (LATE) in which the IV estimates capture the effects for the subpopulation that takes up offers to attend a top *preuniversitario* due to our experiment (the compliers). Using the methodology suggested by Kling (2001), we find that our IV estimates capture mostly effects from the groups with lower *ex ante* probability of attending a top *preuniversitario*. We also show that the monotonicity assumption is likely to be met in our context.

Rigorous evidence on the effect of test preparation is limited. Numerous studies have highlighted correlations between test preparation and increases in test performance, most of them based on SAT preparation in the United States (Becker 1990; Powers and Rock 1999; Briggs 2001; Buchmann, Congdron and Roscigno 2010). However, much of this evidence suffers from large selection bias. The most rigorous study we have found within this literature is an RCT by Allalouf and Ben-Shahkar (1998) implemented in Israel, for which results indicate that test preparation results in small, yet positive and statistically significant test scores gain, although their study suffers from large attrition, which allows them to present only treatment on the treated effects.

Additionally, this is not the first study that tries to answer our questions in the Chilean context. Núñez and Millán (2002) found large gains from test preparation for low-income students. However, this study was conducted in a laboratory setting, using a remarkably small sample, and suffers from certain methodological weaknesses. In this respect, our paper presents some of the first rigorous evidence on the direct effect of test preparation on test scores in Chile and, to the best of our knowledge, in other emerging and developing countries. Moreover, our paper differs from the current literature in several ways, by providing novel evidence on the implications of obtaining higher standard test scores for access to higher education. Additionally, we expand on the focus of the current literature on the US and Israel, which may have limited applicability to educational systems in developing and emerging countries.

In another respect, we also contribute to the literature on how people change their actions and beliefs after they receive a positive shock, such as receiving an offer of a scholarship.[2] This could affect the updating of beliefs about one´s own social mobility

[2]Friedman *et al.* (2011) also show how educational opportunities induced by RCTs change policy actions and beliefs.

(Piketty 1995) or affect beliefs about the degree of fairness in society (Alesina and Angeletos 2005). These beliefs, in turn, may affect policy choices (e.g. in terms of redistribution). In fact, according to our results, being offered a scholarship significantly changed the behavior of students related to the educational policy discussion.

In a first set of findings, students who were offered scholarships in the second cohort supported student protests less strongly and had a higher probability of graduating from 12th grade. It is worth mentioning that in 2011 the Ministry of Education (MINEDUC) set up a free exam system that excluded students from schools susceptible to takeovers (i.e. schools occupied by student protesters), which could have led to the students losing a school year. As a way of protesting, many students chose to repeat their grades and not take these exams. The scholarship offer decreased the number of students repeating 12th grade as a mode of protest.

At the same time, in a follow-up survey we asked students how they would allocate US$20 million of public money for several uses in education, including reform of the higher education admission system, public expenditure on higher education, investing in the fiscalization of higher education institutions, investing in decreasing fees for higher education, and improving the quality of the higher education system. Our findings indicate that students from low-quality schools in the 2010 cohort would spend about US$0.3 million more on improving the higher education admission system and about US$0.3 million less on decreasing fees for higher education institutions.

These results are magnified when one considers the effect of attending a top test preparation institution. These results are concentrated on students from low-performing schools in the first cohort, who greatly increased their access to PSU preparation thanks to the scholarship program, but who did not see any significant improvement in their outcomes. The results disappeared over the second year, when students improved their performance and their access to higher education. These results are interesting in themselves and suggest that the opportunities that high school and university students receive affect education policy.

In addition to presenting evidence on the questions discussed above, this paper also presents evidence from a peer effects intervention that was implemented within the test-preparation institution we worked with. This second study included in this paper is related to an expansive literature on peer effects in the classroom (Hoxby 2000; Sacerdote 2001; Angrist and Lang 2004; Winston and Zimmerman 2004; Duflo, Dupas and Kremer 2011; Lavy, Paserman and Schlosser 2012; Pop-Eleches and Urquiola 2013), and contributes to it by adding evidence from an out-of-school setting, where circumstances may differ from those in school settings.

Within the peer effects literature, mixed evidence favors two potential effects. The first is that students in groups with large concentrations of relatively high-ability (low-ability) students will be benefited (harmed) from the characteristics of their peers through a spillover effect. Hoxby (2000) explores this hypothesis and finds evidence in favor of it. The second effect builds on the fact that a reduction in the variance of peers' characteristics might influence teachers' ability to be effective and reduce class disruption. Duflo, Dupas and Kremer (2011) explore this tracking hypothesis through an RCT in Kenyan schools, finding strong evidence in favor of the hypothesis, and providing evidence against the historic view of linear-in-means peer effects.

Concretely, we implement a second RCT by which a subsample of students are randomly assigned to tracked or non-tracked classes (i.e. classes that do or do not separate students by academic ability), in the spirit of the Duflo, Dupas and Kremer (2011) study. Our results show that the impact of tracking was null on language performance, and more positive but not statistically significant in math and science. Although this lack of significant impacts may be driven by the small sample size with which we performed these estimations, we also hypothesize that the traditional school set-up in which students share one classroom for an extended period of time might differ from the set-up in which we work, where attendance is voluntary and interaction is limited to only a few hours per week.

While Duflo, Dupas and Kremer (2011) argue that impacts of tracking operate by allowing teachers to adapt their teaching more accurately to a less varied student audience, we do not find systematic evidence of the impact of tracking on a number of teaching quality and class interaction outcomes gathered through class observations. In all, these results suggest that the underlying logic of tracking arguments relating to changes in teacher practices may not apply to settings different from schools. More research is needed to understand whether this is the case.

The rest of the paper is organized as follows:

- Section 2 provides background on the Chilean higher education system.
- Section 3 discusses the motivations for the protests that occurred in 2011, and describes how they affected the results from this intervention.
- Section 4 describes the details of the scholarship program and the tracking program structures.
- Section 5 expands on the experimental design, data collection, and statistical methods implemented for our estimations.
- Section 6 presents descriptive statistics.
- Section 7 presents results from the scholarship experiment, discussing the effect of test preparation on test scores, access to higher education, and other outcomes.
- Section 8 presents results from the tracking experiment.
- Finally, Section 9 concludes and discusses policy implications from our results.

2 Higher education in Chile

After completing secondary school, students choose from three forms of higher education: *Centros de Formación Técnica* (Technical Education Centers) (CFT), which offer two-year technical degrees; *Institutos Profesionales* (Professional Institutes) (IP), which offer four-year technical-professional degrees; and Universities (U), which offer five-year degrees, most closely comparable to traditional bachelor's degrees in the US and other developed countries. Universities are usually recognized as the highest in average quality, and receive roughly 48 per cent of all degree-seeking students (CNED 2012).

Chilean universities are classified as either traditional or private; the 25 universities (public and private) founded before 1981 constitute traditional universities, while those established thereafter constitute private universities.[3] While there are many strong private universities, those in the traditional sector are generally recognized as being of higher quality than those in the private sector. However, there is substantial heterogeneity in terms of suppliers' quality in both sectors (Reyes, Rodríguez and Urzúa 2013).[4] Traditional universities also hold a common membership within the Chilean Universities Presidents Council (Consejo de Rectores de las Universidades Chilenas, henceforth CRUCH), an administrative body that manages and standardizes their admissions processes and also includes some of the top private universities.

The higher education admissions process is principally based on a standardized test, the PSU, which is designed primarily as a test on the contents of the government curriculum for high schools. The test takes place once a year, at the end of the school year. It is composed of two mandatory sections, language and mathematics, as well as two elective sections, history and science. Each section is scored 150–850, normalized to a mean of 500 with a standard deviation of 110. The test is set over two days and typically takes eight hours to complete. The scores can be used in two admission processes.

CRUCH has administered the PSU since 2003 as the primary admission component to access traditional universities. In fact, a combined weighted average of the PSU score and high-school grade point average (GPA) – where PSU sections considered and weights vary across careers – determines admission to all traditional and several private universities; no other factors are considered.[5] Although it is not mandatory, almost all universities, IPs and CFTs in Chile require the PSU for admission, and minimum PSU scores are often required to receive student loans. For these reasons, almost all high school seniors choose to take the PSU, including 97.6 per cent of the students in our study.

[3] Reyes, Rodríguez and Urzúa (2013) and Figueroa and Huepe (2013) present detailed descriptions of the higher education sector in Chile.

[4] For instance, only traditional universities ranked among the top 50 universities in the most recent QS Latin American University Rankings, but at the same time the worst Chilean university included in the ranking is a traditional university. In contrast, the first private university is ranked in 52nd place.

[5] Affirmative action programs at the top universities cover a small share of enrollment, but these schemes are also closely related to PSU scores and high school grades.

In a set-up like this, performing well in the PSU becomes a key determinant of the higher education outcomes of each student. For instance, Hastings, Neilson and Zimmerman (2013) exploit the discontinuities created by admissions cut-offs for different top programs to estimate the effects of entering a more selective track on labor market outcomes. They find that returns are positive and significant only among more selective programs, and also that there are no differential outcomes for students from different socioeconomic backgrounds admitted to these degrees.

This explains why, to prepare for the PSU, high school seniors often enroll in *preuniversitarios*, educational institutions specifically designed to improve performance in the PSU exam. *Preuniversitarios* typically complement students' final high school year, offering weeknight and weekend classes corresponding to each of the students' test subject areas. Cost and quality vary widely among *preuniversitarios*, ranging from free non-profit institutions to expensive private educational enterprises. As previously discussed, high-quality *preuniversitarios* typically charge over $1,500 per year for their full range of test preparation services.

These high tuition costs may impose a barrier for low-income students to attend these institutions. While schools prepare students to an extent, preparation is in general of substantially lower intensity than at *preuniversitarios*. Other *preuniversitarios* exist in different formats (and with different prices): from online *preuniversitarios* to *preuniversitarios* taught by volunteers, passing from private non-top *preuniversitarios* to test preparation with individual tutors. There is no public information or regulations regarding the quality of these suppliers.

Access to higher education is highly unequal in Chile: as CASEN (2011) shows, while the higher education enrollment rate for young people between 18 and 24 years of age is 63 per cent in the highest income decile, it is only 22 per cent in the lowest income decile. Some of this sorting may be attributed to pre-existing sorting at primary and secondary education levels (Hsieh and Urquiola 2006) and to the influences of home environment (Reyes, Rodríguez and Urzúa 2013). However, *preuniversitarios* might also be playing a role in this sorting and, if anything, increasing such sorting by providing additional PSU preparation to high-income students.

3 Students' protests

In 2011 Chile experienced a series of massive student-led protests. The protests occurred in response to widespread discontent with the course of the Chilean educational system, alleging large inequalities in access and quality. The movement brought together secondary and university students, who were requesting structural changes to the system. These manifestations of protest were the greatest and longest in Chilean history. Initial demands included increasing spending on education, higher education regulation, and changes to the higher education admissions system, among others.[6] As the movement got stronger, claims focused on the idea of restoring the

[6] The main objective of the movement was to "build an educational project that *[contd. on p.9]* is constitutionally guaranteed as a universal social law in all its levels, that is founded on a system of education that is public, democratic, pluralistic, free and of high quality, oriented to the production of knowledge for a comprehensive and equitable development and to meet the

public education system that included the demand for free access to high-quality education and the prohibition of for-profit schools and higher education institutions.

Student protests began on May 12, with the first march in Santiago gathering more than 15,000 students.[7] Throughout June, higher education and secondary students respectively occupied multiple universities and schools. On June 11, students announced a radicalization of the students' movement; the march that followed on June 16 gathered more than 80,000 people in Santiago. Over the following months, multiple marches took place in Santiago, gathering up to 100,000 people. According to the students' movement, the number reached 200,000 on certain occasions.

The marches that took place in 2011, as well as the school takeovers, resulted in a significant loss of school days. Students in our sample lost on average 37 school days due to student protests in 2011, with some losing up to 101 school days. The movement was particularly strong among public and private subsidized schools of higher quality; in fact, while students attending low-quality schools in our sample lost on average 20 school days due to student protests, students attending high-quality schools lost on average 52 school days.

Given the significant loss of school days, in August 2011 the government launched a program called Save the School Year (Salvemos el año escolar). The program offered different solutions, including taking classes at alternative institutions (because student protesters continued to occupy schools), and free out-of-school examinations. Many students rejected the government's plan and chose to repeat the school year as a form of protest.

Due to the protests, the number of students taking the PSU test dropped significantly. Figure 2 describes the drop in the ratio of the number of students taking the exam to the total number of students. Whereas in 2010 a total of 250,758 students took the PSU exam, in 2011 this figure dropped to 231,170 (DEMRE 2012). Furthermore, the protests not only impacted the number of students taking the test but also the pool of students doing so. According to our data, dropout rates were particularly high among students attending low-performing schools (Figure 3), which is particularly relevant for our study because the PSU provides an ordinal classification of students.

Overall, the protests that took place on 2011 had a direct impact on our results, and we believe that they might be the reason behind the heterogeneous effects found in this study. We will return to this point later. Figure 1 shows a timeline of the events, describing how protests overlapped research and intervention activities.

---

needs of Chile and its people" (CONFECH 2011). The specific demands were: (i) increases in government expenditure on education; (ii) explicit rights for students and administrative staff to participate in the government of universities; (iii) equality in access to higher education; (iv) a reform of the scholarship system in the short term; (v) banning private banks from participating in the provision of higher education loans; (vi) ending for-profit education institutions; and (vii) a system of public and free education in the long run. For more details, see CONFECH (2011).

[7] The size of and support for the riots were mostly unprecedented before this period.

Figure 1: Timeline

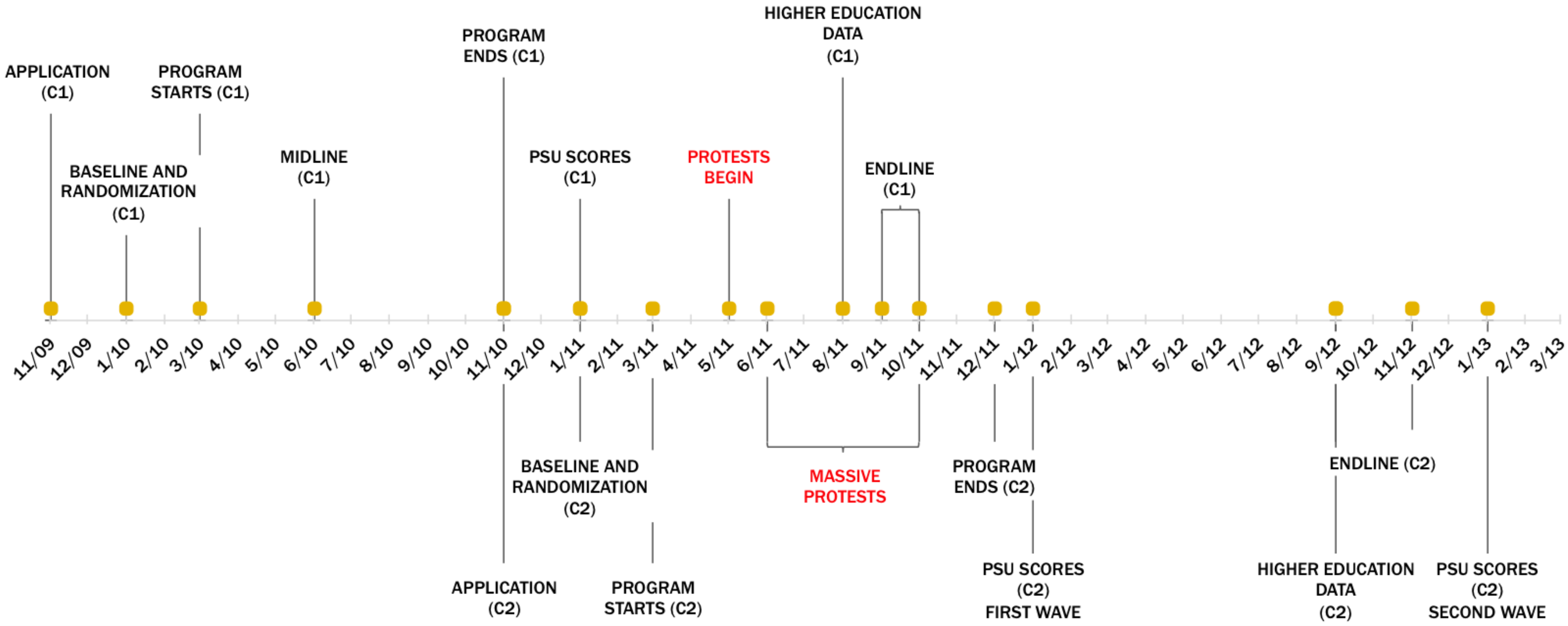
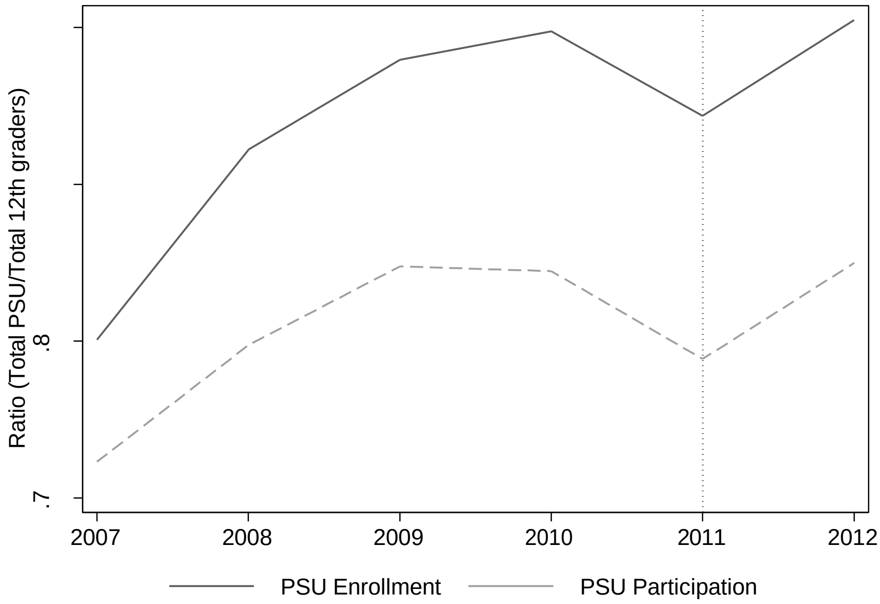
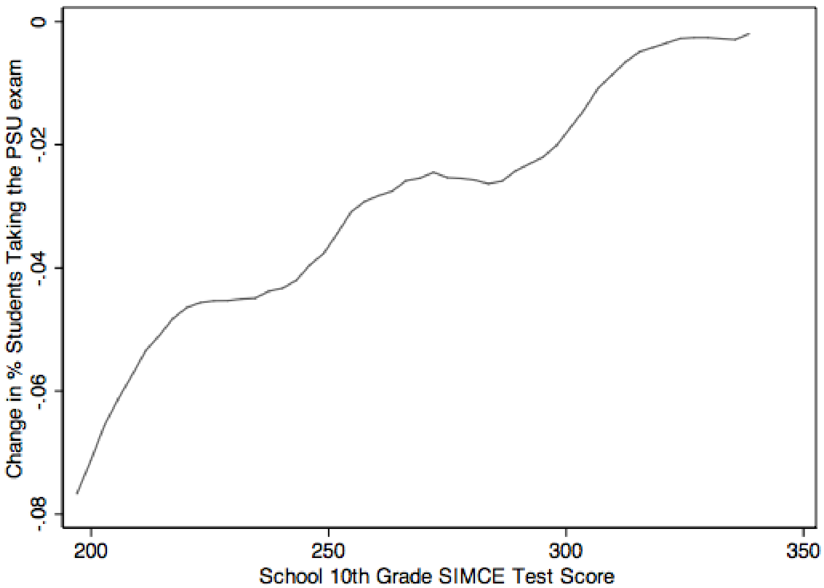


Figure 2: PSU enrollment and participation



Note: This figure shows the ratio of students enrolled and participating in the PSU over the total number of 12th graders. Only students from the promotion of the year of the PSU are considered. Enrollment refers to the number of students registered to take the PSU. Participation is defined as the number of students who actually took the exam.

Figure 3: Percentage of students taking the test and high school performance, 2011 vs. 2010



Note: This figure shows a kernel regression of the change in the percentage of students taking the PSU exam between 2010 and 2011, and SIMCE scores. Data is at school level. A rule of thumb bandwidth estimator is used. Only local mean smoothing is included.

4 The programs: scholarships and tracking

During 2010 and 2011, our team and Preu UC offered a scholarship program to attend Preu UC, one of the top three *preuniversitarios*, located in one of the country's premier universities, the Pontifical Catholic University of Chile. The program was designed to attract above-average students from schools with a high proportion of students from low- and middle-income families who would potentially otherwise not be able to attend a high-quality *preuniversitario*. To be eligible for the scholarship, students needed to fulfill three requirements: (1) hold a GPA of at least 5.5 out of a possible 7.0 (the mean GPA of students in Chile is 5.4 with a median GPA of 5.4)[8]; (2) live in Greater Santiago, the region where Preu UC is located; and (3) be entering 12th grade, which is their fourth and final year of secondary school, at a public or subsidized private school.

Scholarship students received a 65 per cent discount, from an annual cost of approximately $2,000, on *preuniversitario* services, including enrollment in three courses and five full-length practice tests.[9] Considering that the families in our sample spent on average $400 a year per student on public and private voucher schools, the cost of attending a top *preuniversitario* is remarkably high, and clearly unaffordable for low- and middle-income families in Chile.

Classes for Preu UC began in March of each year, and lasted through early December, corresponding roughly to the academic year. Scholarship students enrolled on three courses: one language course, one math course, and one additional course, either history or science. Students who opted for the science course also took an additional science elective – biology, chemistry, or physics – corresponding with the science elective component of the PSU science section. Language, math, and history classes each met twice a week, with classes of approximately one hour and 20 minutes. Science classes met four times a week, with classes of the same duration.

As shown in Table 3, a total of 2,077 eligible students applied for the scholarship. It should be noted that students who chose to apply for the program were high-performing and attended institutions with relatively high academic outcomes. In fact, most of the applicants attended middle (43%) and middle-high (40%) socioeconomic status schools, and a smaller percentage of them attended middle-low, low, and high socioeconomic status schools (17%).[10] More detailed information can be found in Table 5, which compares the socioeconomic composition of students in our sample to the socioeconomic composition of eligible students (i.e. students attending public and private subsidized schools and following a non-vocational track) in Chile and Greater

---

[8] Data on school performance for 11th grade students attending public and private subsidized schools in 2011.

[9] The total cost of the language, math, and science elective course package offered by Preu UC was approximately $2,500 in 2011. For students enrolling in the history class instead of science, the package cost was $1,600.

[10] MINEDUC implements the schools' socioeconomic classification based on (i) years of education of parents; (ii) self-reported household income; and (iii) number of vulnerable students. For more details see MINEDUC (2012).

Santiago. Table 5 shows that the experiment's sample is skewed toward middle and middle-high socioeconomic status institutions, which leads to underrepresentation of students from institutions with low, middle-low and high socioeconomic profiles.

The students in our sample are also characterized as attending schools with above average academic performance; according to our data, students attended schools with an average SIMCE score[11] of 284 points, which corresponds approximately to the 75th percentile of the distribution of the SIMCE results. Furthermore, students who applied for the program had an average GPA of 6.0, which puts them approximately in the 75th percentile of the distribution of academic performance.

It should be noted that the composition of students differs between cohorts. On the one hand, the number of applicants decreased by 16 per cent in the 2011 cohort. On the other, the data suggest that the pool of applicants also changed in terms of socioeconomic and other characteristics (see Panel A in Table 6): students of the second cohort in the study went to better high schools, had higher expectations in terms of their PSU scores, and had slightly better high school performance and attendance rates (all significant at the 1% level).

However, it should be noted that the differences observed in terms of socioeconomic variables between cohorts are not the ones driving the differences observed in treatment effects across cohorts. In some exercises (Appendix Table A1), we analyzed the interaction between the treatment and baseline variables only to find that they cannot explain the differences in treatment effects across cohorts.

One important change across years that can explain the differences observed between cohorts is the existence of changes that occurred in the diffusion process between years. The scholarship program was initially advertised on radio, television, in local newspapers, at local high schools, and online. However, in the second year, the program was advertised only in newspapers and at local high schools, the diffusion process being therefore remarkably less than for 2010. Also, that the scholarship was at a 65 per cent discount rather than a full tuition discount was made more explicit for 2011 to avoid payment misunderstandings and difficulties registered with our partner institution during 2010.

Of the 2,077 students who applied for the scholarship, 1,266 were randomly selected to receive an offer. Recipients were notified in January of each year, with matriculations offered throughout February. Ultimately, 908 students accepted the offer (Table 3). Acceptance rates also differed between cohorts, as columns (1) and (2) in Table 11 show; whereas the scholarship acceptance rate for students in the first cohort was 77.2 per cent, it was only 64.4 per cent for the second cohort. This possibly reflects that the latter had better alternatives than the former in terms of test preparation.

[11] SIMCE is an acronym for Educational Quality Measurement System (Sistema de Medicíon de Calidad de la Educación), a nationwide exam for students in grades that change between years, containing at least language and math sections. These are the scores we use across the study to account for high school quality.

To better understand the differences in acceptance rates between cohorts, one has to take a closer look at the private market for *preuniversitarios*. In Chile, the market for *preuniversitarios* is very active, and private suppliers face a high level of competition. To differentiate themselves in terms of quality, massive *preuniversitarios* have developed marketing strategies – using television, radio, and newspaper ads –that emphasize the number of former students performing well in the PSU.

This explains why *preuniversitarios* offer scholarships to high-performing students attending high-quality schools, hoping that these students can outperform on the PSU. In fact, about 50 per cent of students in our experiment who were not offered a scholarship attended a top test preparation institution; 40 per cent were students attending low-quality schools, and 61 per cent were students attending high-quality schools. We believe that most of these students did so by receiving at least partial scholarships.

A second important change that took place across the years is related to changes in the amount and timing of scholarship offers by other *preuniversitarios*: unlike in previous years, Preu UC's top competitors (CEPECH and Pedro de Valdivia) released their scholarship offers before us on 2011. With respect to the number of scholarships handed out, our data indicate that the takeup of top *preuniversitarios* among members of the control group increased from 41.0 per cent in the first cohort to 63.3 per cent in the second cohort. This difference is also apparent in students from both low- and high-quality schools. Actually, takeup of top *preuniversitarios* increased from 33 per cent to 52 per cent among students from low-quality schools, and from 53 per cent to 73 per cent among students from high-quality schools.

We do not know the exact percentage of students who received scholarships, but Table 2 presents information from our follow-up survey in terms of the expenditure of students in the control group who attended top *preuniversitarios*. This table shows that the median student from a low-quality school who attended a top *preuniversitario* decreased his or her nominal expenditure by about 28.9 per cent (31.3% in real terms) between 2010 and 2011. In contrast, the median student from a high-performing school decreased his or her nominal payments by just 6.2 per cent (9.2% in real terms).

The table also presents information for students paying 0, less than US$600, and less than US$1,000 (which correspond to two thresholds that imply relevant discounts from list prices). The data suggest that there are significant increases in the proportion of students attending top *preuniversitarios* with total or partial scholarships, especially from low-performing schools. In sum, this evidence suggests that the market became much more active in 2011, in particular offering discounts and scholarships to students from low-performing schools.

Of the 908 students who accepted the scholarship offer, 840 (92.5%) completed the program. Those who withdrew from the program were not charged for the months they did not attend. The most common reasons for attrition were stress, inability to pay, and school and work commitments. Notably, there are also relevant differences in completion rates between both cohorts of students. Whereas 88.1 per cent of students in the 2010 cohort completed the program, as many as 98.1 per cent did so in the 2011 cohort. Overall, the probability of completing the program among students who were offered a scholarship is non-statistically different between cohorts, because students in the 2011 cohort had fewer chances of accepting the offer; but once accepted, they had a higher chance of completing the program.

Once enrolled, students began attending classes at Preu UC. To facilitate the study design, Preu UC separated the classes involved in this experiment from those that were not involved. Therefore, scholarship students took classes strictly with other scholarship students. Among scholarship classes, a number of students were assigned to classes according to their ability, with the objective of examining the presence of peer effects.

This implied that within the set of scholarship students' classes, there were a number of low-, high-, and mixed-ability tracked classes, with assignment to them based on students' baseline PSU scores. Approximately two-thirds of all math, language and science sections were designed as tracked classes; history classes were not included due to logistical limitations in the quantity of classes offered. Low- and high-ability classes were composed exclusively of students with these ability levels. Mixed-ability classes were constructed to evenly balance high- and low-ability students.

Neither students nor professors were notified of these arrangements, and classes were administered identically to non-tracked classes.[12] As shown in Table 4, a total of 559 students were assigned to tracked language classes, while 574 and 371 were assigned to tracked math and science, respectively. Distribution of students across different types of tracked classes was almost uniform.

5 Experimental design

5.1 Scholarships assignment

The scholarship eligibility requirements mentioned in the previous section were set to ensure that the sample consisted of above-average high-school students from poor and middle-income families. In particular, a GPA of 5.5 is a signal of the former, while enrollment in a public or private subsidized school is a noisy proxy for the latter. The requirement that applicants lived in the Metropolitan Region was simply used as a practical device to limit drop-outs from the program, which would be more likely in cases in which transportation costs were higher.

Once samples of eligible applicants were collected, baseline exams and a questionnaire were applied to them. Using data gathered from those instruments, students were randomly assigned to treatment and control groups, where the treatment is defined as being offered a scholarship for Preu UC. To ensure balance between

[12] Once assigned to a particular class, students were allowed to change classes only by request, and they were not allowed to move into peer effects classes of ability levels different from their own.

groups in relevant characteristics, this assignment was stratified by high school dependence (i.e. public schools or private subsidized schools), high school quality, gender, baseline test scores, and high school grades. This randomization was performed within each cohort of students and, for both cohorts randomization was performed with a target of obtaining no more than 500 acceptances per year, which was the number of scholarships Preu UC was willing to provide.

Thereafter, a simple randomization was performed among students not offered the scholarship to create an ordered waiting list for the remaining scholarship offers, which were extended only as students initially offered scholarships rejected them. Given that the waiting list was sorted randomly, this process preserved the balance between treatment and control groups in both cohorts.

As presented in Table 3, of 2,077 eligible students who applied for the scholarship, 1,266 were randomly selected to receive scholarship offers: 614 in 2010 and 652 in 2011. A total of 161 students from the waiting list were offered scholarships: 61 in 2010 and 100 in 2011. These students were notified of their scholarship offer one month later than their fellow scholarship recipients and before the actual program started. However, they were otherwise treated exactly the same as their counterparts. Of all participants offered scholarships, 908 accepted: 488 in 2010 and 420 in 2011. This implies an acceptance rate of 71.7 per cent, with the previously mentioned differences between cohorts.

5.2 Tracking assignment

The design of the peer effect experiment follows Duflo, Dupas and Kremer (2011), in the sense that it randomly assigned students to classes where peers were either of high ability, low ability, or a mixture of both. The implementation of this design consisted of two stages. First, the number of classes of each ability level was determined by the number of scholarship students jointly with the preferences of students in terms of desired class time blocks. Thus, tracked classes were created only when a minimum of three classes for the same subject were to be offered within the same time block.

In such cases, three classes were created, one for each of the ability levels mentioned above. When this requirement could not be met, classes were constructed as traditional classes without compositional constraints. In fact, as already discussed, history classes were offered only as traditional classes, because there was no time block where three history classes were offered simultaneously. If more than three classes were offered in the same time block, additional mixed classes were formed. Ultimately, of a total of 73 language, math, and science classes offered over the two years of the study, 46 were designed as tracked classes.

After class types were determined, students were randomly assigned to each class type as follows. Students were listed by baseline PSU score for each subject in each time block with tracked classes. In the case of science classes, the PSU score used was an average of PSU scores in baseline language and math exams.[13] A median PSU score was calculated in each list and used to label students as either low or high ability. Every third student on the list was assigned to the mixed ability class. The remaining students were assigned to either the high- or low-ability class, corresponding to their aforementioned ability level. In this way, classes were initially balanced in their composition between tracked classes and non-tracked ones. As previously mentioned, students were allowed to change classes only if doing so did not conflict with the peer composition type of the classes.

A detailed breakdown of the class types by cohort is provided in Table 4. It shows that sample sizes for this experiment are somewhat small, with only 559, 574, and 371 students, respectively, in language, math, and science. This reduced sample is a consequence of the aforementioned restrictions.

5.3 Data collection

This paper uses data gathered by the research team from both cohorts participating in the program, as well as administrative data. This section describes the relevant data-collection activities and instruments used in the study.

First, baseline data were collected from each student at the time of application to the scholarship program. This included information on individuals' previous educational attainment, educational expectations, plans in terms of PSU preparation, self-perception measures in terms of control over own life, information about familiar background and household characteristics, as well as contact information.

In addition, all applicants were required to take an unofficial baseline PSU exam offered in November and December of the year prior to enrollment at a *preuniversitario*. This allowed us complete baseline data on a number of relevant variables for all 2,077 students in the sample of our study. In what follows, we use this baseline data to check for balance between groups.

Additionally, educational data were collected from MINEDUC, including individual- and school-level characteristics. The former included 11th- and 12th-grade high school GPAs, attendance rates, repetition rates, and information on students' higher education enrollment, including initial institution and career choices. The latter included information on high school dependence (public or private subsidized), high school district, average SIMCE test scores, and the occurrence of the 2011 student protests on school days.

In what follows, we use administrative data on school characteristics and students' prior academic performance to check for balance between groups. Additionally, we use administrative school information to check for heterogeneous impacts on school

[13] Note that only language and math PSU exams were given at baseline, which did not allow us to use a science score as the ranking instrument. However, for students in the control group, the correlation between the average of these two scores and the endline science PSU score was 0.81.

quality. Finally, we use the data with respect to students' 12th-grade performance and higher education enrollment to estimate the impact of the program on students' performance in their final year of school, and on higher education access.

A second source of administrative data on educational outcomes was the Department of Evaluation, Measurement and Educational Enrollment (Departmento de Evaluación, Medición y Registro Educational) (DEMRE), which administers the CRUCH university admissions process. Data collected included individual data on students' PSU performance by area, including participation, scores, number of correct answers, and percentile, all of which allow us to estimate the impact of the program on standardized test performance.

A third source of data was the My Future website (www.mifuturo.cl) run by MINEDUC, which offers data on the average PSU scores and labor market outcomes obtained by students belonging to the institutions and careers that students in our sample decided to study for.[14] Outcomes included are employment rates after the first year in the labor market, and wage ranges after the fourth year in the labor market. Our intention was to estimate program impact on the quality of the careers opted for by students in the sample (measured by indexes that map individual PSU scores to average PSU scores and expected income).[15]

We collected process information on both cohorts in the study through active monitoring of randomly chosen classes. Monitoring visits were conducted from early June to early November in 2010, and from mid-April to early December in 2011. During these visits, we gathered on classroom infrastructure, teacher and student punctuality, classroom management, availability of learning opportunities, teaching techniques and quality, student participation, and use of additional materials.

Monitoring was done by trained psychology students, all of whom were in their final year of study. They observed classes in their entirety, all visits were made without prior notice, and each class in our sample was visited at least once. As previously stated, Preu UC separated the classes involved in this experiment (scholarship classes) from those not involved (non-scholarship classes). To check whether the program operated equally in both cases, we included scholarship and non-scholarship classes in the analysis.

A total of 613 classes were observed, of which 308 corresponded to classes with scholarship students, and 147 were included in the tracking experiment. In what follows, we use this information to check for differences between scholarship and non-scholarship classes, cohort 2010 and cohort 2011 classes, classes included and not included in the tracking experiment, tracking and mixed classes, and high- and low-ability tracking classes.

[14] We collected information for programs representing 68.8% of the students in the sample.
[15] To have a measure of expected wages and employment of students not enrolled in higher education, we use data reported in MINEDUC (2013) for high school graduates for the same cohorts as the ones included in the My Future dataset.

Finally, we collected endline data from study participants through a follow-up survey conducted 10 to 14 months after their cohort's *preuniversitario* end date. Data collection for the first cohort took place in September and October of 2011, overlapping with the student protests, and data collection for the second cohort took place in November 2012 (Figure 1). A professional surveying team carried out the surveys in person, and if this was not possible, then by phone.[16] The team included questions on students' educational and employment status, higher education financing, PSU preparation, *preuniversitario* experience, and educational reform movement opinions. Of the 1,116 study participants in the 2010 cohort, 1,008 (90.3%) completed the survey. Similarly, of the 961 study participants in the 2011 cohort, 867 (90.2%) completed the survey. We discuss the characteristics and implications of attrition rates in Section 6.1.

5.4 Statistical methods

5.4.1 Scholarships program

The random assignment of the treatment across eligible applicants to the scholarships for Preu UC allows us to estimate the effect of offering the scholarships simply by comparing average outcomes of the treatment and the control groups. This estimation allows us to measure the impact of the scholarship on a number of relevant educational outcomes. Additionally, we undertake IV estimations to measure the impact of attending a top *preuniversitario* on the same outcomes.

Regarding the direct impact of being offered a scholarship to Preu UC, we simply run the following ordinary least squares (OLS) regressions to estimate the ITT effect:

$$Y_{ic} = \alpha + \beta T_{ic} + \gamma X_{ic} + E_{ic}$$

where Y_{ic} is the outcome of interest for student i in cohort c, T_{ic} is a dummy variable that equals 1 if the student was offered a scholarship for Preu UC and β measures the impact of the offer.

X_{ic} is a set of control variables at the student and school level, including students' gender, baseline and expected PSU test scores, baseline school grades, self-reported dedication to studying, perception of control over life, propensity to plan, average school SIMCE test scores as a measure of school quality, birth order, maternal education, household income, parental support for education, the number of books at home as a measure of household cultural capital, and dummies for the strata to which treatment was assigned. Given that treatment assignment is random, β should not change when including control variables in the regression; nonetheless, adding them increases the precision of our estimates, which is the reason we do so.

[16] 293 surveys (30%) were conducted on the phone in 2010, and 168 surveys (19%) were conducted on the phone in 2011.

Given that the scholarship program resulted in a significant increase in takeup– where being offered a scholarship increased the likelihood of attending a top test preparation institution by 37 p.p. – we can estimate the impact of attending a top *preuniversitario*. To do so, we estimate the following IV regression:

$$Y_{ic} = \alpha + \beta^{IV} TP_{ic} + \gamma X_{ic} + E_{ic}$$

where all the variables are the same as in equation 5.1, except for TP_{ic}, which is an indicator taking the value 1 if the student attended a top *preuniversitario*. Given that attending a top *preuniversitario* is arguably determined by unobservable characteristics of students or families, this indicator is presumably endogenous, and we instrument it in this regression using the ITT dummy T_{ic} as an IV. β^{IV} will be the estimate for the impact of attending a top *preuniversitario* on our outcomes of interest, and should have the interpretation of a local average treatment effect for the share of students who were impacted by scholarship offers in terms of their attendance at a top *preuniversitario*.

Finally, to explore the existence of heterogeneous effects across cohorts and across levels of school quality, we also estimate regressions including interactions between treatment variables, cohort (C1, C2), and school quality level, the last being defined as of high quality if it is above the median school in the sample in terms of average SIMCE, and of low quality if below it (HQ, LQ). This modifies equation 5.1 and leads to the following specification:

$$Y_{ic} = \alpha + \beta_1 T_{ic} + \beta_2 T_{ic} \cdot C2_{ic} + \beta_3 T_{ic} \cdot HQ_{ic} + \beta_4 T_{ic} \cdot C2_{ic} \cdot HQ_{ic} + \gamma X_{ic} + E_{ic}.$$

In such a specification, the impacts for different groups of students are obtained by adding coefficients. For example, the impact on students from high-quality schools that took part in the program in the 2010 cohort would equal $\beta_1 + \beta_3$, and the impact on students from high-quality schools that took part in the program through the 2011 cohort would equal $\beta_1 + \beta_2 + \beta_3 + \beta_4$.

In the case of IV estimations, we modify equation 5.2 analogously by including the same interactions between TP, $C2$, and HQ. We instrument these interactions using the interaction between the treatment assignment variable, T_{ic}, and the interactive variables, which leaves us with the same number of endogenous variables as instruments. At the bottom of all tables we include information about the statistical significance of the relevant effects that correspond to the sum of coefficients.

5.4.2 Tracking program

Regarding the tracking experiment, our empirical strategy closely follows the one that Duflo, Dupas and Kremer (2011) implemented, in terms of exploiting random assignment of tracking to students to identify its impact. Along these lines, we estimate the following regression:

$$Y_{ic} = \alpha + \beta TR_{ic} + \gamma X_{ic} + E_{ic}$$

where Y_{ic} are PSU test scores for student *i* in cohort *c*, TR_{ic} is an indicator that takes the value 1 if the student was assigned to a tracked class, β is a measure of the impact of tracking on PSU test performance, and X_{ic} is a set of control variables, in which we include the same variables used for the scholarship experiment's estimation and dummies for time blocks in which students were assigned to tracked or non-tracked classes. We run this regression by PSU area.

Next, among students allocated to the tracking classes, we estimate whether there was a differential effect for the marginal students of being located in a high- or a low-achievement class (Duflo, Dupas and Kremer 2011; Pop-Eleches and Urquiola 2013). We do this by exploiting the discontinuous allocation of students across high- and low-ability classes and implement regression discontinuity design (RDD) estimates (using the Imbens and Kalyanaraman [2012] procedure to compute the optimal bandwidth).[17]

RDD models estimate local linear regression models on both sides of the cut-off using a triangle kernel. In this particular case, the cut-off equals the median baseline PSU score; students above the median access a high-achievement class and students below the median access a low-achievement class. The main assumption is that the marginal students who barely received the treatment in the high-achievement class are comparable to those who barely received the treatment in the low-achievement class.

Finally, given that the tracking experiment affected the mean and variance of the quality of the peers in each classroom, we implement an IV estimation procedure in which the average and the standard deviation of the performance of students in the baseline in each class are treated as endogenous variables, and the random allocation to the mixed classes and the dummy for being allocated to a high-quality class (a quasi-random allocation based on a discontinuity) are used as IVs. We also control for cubic functions of the baseline of each test in each estimation and include a vector of other baseline variables.

6 Descriptive statistics

6.1 Balance between groups

To study the validity of random assignment of both experiments, we test for differences between treatment and control groups in a group of baseline variables. Moreover, we test for differences in baseline variables among students assigned to tracking and non-tracking classes. Finally, we test for differences between attriters and non-attriters (i.e. students who did or did not complete our follow-up survey), checking for differences between attriters assigned to the control and treatment groups.

Regarding the scholarship experiment, we first test for differences between students who were offered scholarships and those who were not. The results are shown in

[17] As a robustness check, we also estimated the RDD models using the Calonico, Cattaneo and Titiunik (2014) procedure, which produced very similar results.

Panel B of Table 6. No statistically significant differences are found in terms of students' gender, high school quality, previous high school performance and attendance, baseline math test scores and PSU test scores expectations, attitudes related to PSU preparation, our index of self-control, our indicator of propensity to plan, and the household-related variables we are using in the analysis.

Baseline language test score results seem to be the only variable for which there are statistically significant differences between groups, for which treated students performed 6 points (0.08σ) higher than students in the control group (significant at the 10% level). Note that because we use these variables as controls in our regressions, any differences between groups would be controlled for in our analyses.[18]

We also test for differences between groups in the tracking experiment. In this context, we test for differences between students assigned to tracked and non-tracked classes by subject (language, math, and science). Panels A, B, and C in Table 8 present the results for students included in the tracking experiment in language, math, and science, respectively. The results show again that, in general, random assignment to tracked classes created groups that have almost no statistically significant differences in terms of the variables considered for this analysis.

For instance, students assigned to tracked classes for language differ from those assigned to non-tracked classes only in having 10.7 per cent more female students in the group (significant at the 5% level) and in having a slightly lower attendance through 11th grade (significant at the 10% level). Students assigned to tracked classes for math have lower expected PSU scores for language and math by 14.5 points (0.18σ) and 15.4 points (0.17σ) (significant at the 10% and 5% levels, respectively). Finally, no statistically significant differences are observed between students assigned to tracked and non-tracked science classes.

As mentioned before, efforts were made to reduce attrition from the survey, which left the attrition rate at 9.7 per cent. In this regard, we first test for differences between attriters and non attriters, for which Panel A in Table 9 shows the results. These tests show just a few differences between those groups and if anything it seems that attriters were relatively more motivated, reflected by positive statistically significant differences in 11th-grade GPA (significant at the 5% level) and intention to attend *preuniversitario* at baseline (significant at the 1% level). They also came from better-off households, as measured by our asset index and by the number of books at home (both significant at the 1% level).

Next, we compare attriters assigned to the treatment and control groups. This comparison is particularly relevant because differences between these two groups would constitute an effective threat for our identification strategy provided by random assignment. The results for these tests are shown in Panel B in Table 9, and show that attriters in both groups are balanced in most characteristics, the only difference

[18] We performed the same balance tests in each cohort in the study, with the result that groups are balanced in almost all variables tested for in both cohorts. Results are available from the authors on request.

being their attendance at a *preuniversitario* during 11th grade (significant at the 10% level). This result rules out the possibility that sample attrition could be a problem for our empirical analysis, at least regarding observable characteristics at baseline.

6.2 Class observations

As mentioned earlier, we implemented a program of monitoring activities that consisted of class observations. All classes in our sample were observed at least once, adding up to a sample of 308 observations, 196 in 2010 and 112 in 2011. Using this data, we estimate if there were relevant differences in implementation between different sets of classes using the following estimating equation:

$$Q_{sac} = \alpha + \beta G_{sc} + \gamma X_a + E_{sac}$$

where Q_{sac} is one of a number of quality indicators taken from session *s*, which offered classes in area *a* during cohort *c*. G_{sc} is an indicator that takes a value of 1 if sessions corresponded to a specific group *G* (e.g. a scholarship versus a regular class, a class in the first versus the second cohort, a class in the tracking experiment versus a class not included in that experiment, a tracked versus a mixed class, a high-quality versus low-quality class among tracked classes). X_a includes indicators for the area to which the session corresponded and a dummy for the cohort of the session, and E_{sac} is an error term. Our coefficient of interest in this regression is β, which will measure average differences between groups in terms of each quality indicator.

Results from these estimations are shown in Table 10. The first column presents the average for each variable for the group of classes incorporated in our experiment. Next, we perform comparisons between different groups of sessions. First, in Panel A we compare the classes included in the experiment with regular (non-scholarship) classes offered by this *preuniversitario*. The idea is to check if the treatment we are evaluating corresponds to a regular version of a *preuniversitario*. This is important for external validity, because the students in the experiment are significantly different from regular students (they are poorer and come from different schools).

While all (regular and scholarship) classes are supposed to have the same maximum class size, effective attendance at scholarship classes is higher than at regular classes, scholarship classes having on average 2.4 additional students present (significant at the 1% level).[19] Results imply that in most dimensions, we do not find significant differences between regular and scholarship classes. The small differences observed in students' participation and teachers' control over the class suggest that the scholarship classes actually worked slightly better than the regular ones, despite the bigger class size of the scholarship sessions, with students in scholarship classes answering more questions (0.15σ significant at the 10% level) and teachers increasing their control over class (0.07σ significant at the 5% level).

[19] Personal conversations with our partner organization suggest that this is because scholarship students were less likely to be absent from classes. This is because one of the conditions for keeping the scholarship was not to miss more than three sessions without a proper justification.

When comparing sessions across cohorts, we observe significant differences between the two groups (Panel B of Table 10). On the one hand, in the second cohort, teachers were more likely to adapt contents to students' needs in 0.32σ (significant at the 1% level), students' attitude in class was 0.22σ better (significant at the 5% level), teachers were more likely to provide worksheets in 0.25σ (significant at the 5% level), and they used the projector more often, 0.14σ (significant at the 10% level). On the other, in the second cohort, classes had on average 3.2 additional students present (significant at the 1% level), students were less likely to ask questions in 0.20σ (significant at the 10% level), teachers were less likely to answer questions in 0.26σ (significant at the 5% level), the student–teacher relationship was worse in 0.29σ (significant at the 5% level), and teachers were less likely to refer to worksheets in 0.75σ (significant at the 1% level).

Although there are significant differences between cohorts, there is no clear evidence to indicate that classes worked better one year compared to the other. This is important because, while we find differences in outcomes across cohorts, it is hard to argue that the differences that we observe are a consequence of an improvement in how the program was implemented.

Next, we compare the classes included in the tracking experiment with the other scholarship sessions in Panel C. Again, we find some cases with significant differences: the sessions included in the tracking experiment had on average 1.81 additional students (significant at the 10% level), teachers were less likely to be punctual in 0.61σ (significant at the 1% level), teachers were less likely to understand the contents in 0.28σ (significant at the 5% level), but at the same time the teacher–student relationship seemed to be better in 0.39σ (significant at the 1% level).

Finally, we study whether there were differences among mixed and tracked classes (Panel D), and among high-quality and low-quality classes among the tracked classes. These comparisons are important to understand the potential effects of the tracking experiment. Considering the small sample size used for these regressions,[20] we are interested in reading their results mostly as suggestive rather than as quantitative evidence. Although there are certain differences between tracked and mixed classes, they are remarkably small and, overall, these regressions show that for almost all the outcomes considered, there was no statistically significant difference between tracked and non-tracked classes in either cohort of the study. The only statistically significant differences we find imply that students in tracked classes are less likely to ask questions and that teachers are less likely to use a projector in class.

Next, we compare high- and low-quality classrooms (with an even smaller sample size) in Panel E. The results indicate that there are more variables that are statistically different in this comparison: high-quality classes are larger because students are less

[20] The small sample size is because the unit of observation used for these regressions is the classroom. While a total of 308 classes were observed in the context of the monitoring activities we performed throughout the project, only 147 were done in classes included *[contd. on p.23]* in the tracking experiment.

likely to be absent, teachers are less likely to understand the materials and to talk about students' mistakes. At the same time, the probability of students disrespecting teachers increases significantly. A more positive result is that teachers are more likely to use projectors. In all, we do not have strong evidence suggesting that one type of class worked better than the other. If anything, evidence suggests that low-quality classes worked better.

In all, results of comparison across different groups of classes show a number of statistically significant differences in how the program was implemented, but we do not find systematic and clear patterns in favor of some groups, so it is difficult to attribute impact heterogeneity to these differences.

7 Results from the scholarships experiment

In this section, we present the main results from the evaluation. We begin by estimating the impact of offering the scholarship on entry at a *preuniversitario*. Next, we present the results of the impact of being offered the scholarship on 12th grade performance, test performance, access to higher education, and other related outcomes. We do so by using a simple OLS regression as described in Section 5.4. In each case, we test for heterogeneous effects in terms of school quality and cohorts.

When interpreting the results, it is important to keep in mind that students have outside options for test preparation. First, they can prepare the PSU using less expensive technologies (e.g. school preparation, preparation in free online *preuniversitarios*, and enrollment in non-top *preuniversitarios*). Second, other top *preuniversitarios* also offer scholarships. For instance, among students who rejected the scholarship offer, 68 per cent were attending a top *preuniversitario*, and of these, 17 per cent were studying with full scholarships and about 68 per cent were studying with a significant partial scholarship (see Table 2). Therefore, ITT estimates should be interpreted as the impact of being offered a scholarship to attend a specific high-quality test preparation institution given the other alternatives in place.

Next, given that the scholarship offer had a significant impact on the probability of attending a top test preparation institution, we use this as an instrument for test preparation and include estimates of the impact of attending a top test-preparation institution on 12th grade performance, test performance, and access to higher education (see Section 5.4). It is worth restating that the IV strategy only allows us to estimate a local average treatment effect, which is the impact of attending a top test preparation institution on the subsample of students who attended because they were given the scholarship to attend Preu UC (compliers). We present exercises following Kling (2001) to identify the groups of students most likely to be affected by the treatment.

7.1 Entry to a *preuniversitario*

We start by estimating the impact of offering scholarships for Preu UC on receiving such scholarships (i.e. accepting the offer), on completing the Preu UC program, and, relevantly, on attending a top *preuniversitario*.[21] Results in Table 11 imply that offering scholarships strongly impacted these three outcomes, but also that it did so differently for students from different study cohorts and from low- and high-quality schools.

For instance, the scholarship acceptance rate was 70.7 per cent across the whole sample, as shown in column 1, but was significantly higher for students in the 2010 cohort (77.2%) than for those in the 2011 cohort (64.4%). As discussed previously, these differences are probably a consequence of two facts: (i) that the latter cohort was composed of students who were better off in a number of dimensions than the former, as shown by Panel A in Table 6; and (ii) that other top *preuniversitarios* significantly increased their entry in the second period.

Then, as expected, and consistent with the evidence presented in Table 2, a similar pattern appears in column 3 when differentiating impacts according to students' school quality, with those from low-quality schools accepting the scholarship offer more frequently (75.2%) than those from high-quality schools (66.6%). This combination of results implies that the group that most frequently accepted the scholarship program was the one formed by students in the 2010 cohort who came from low-quality high schools (80.2%), while the group that did so less frequently was the one formed by students in the 2011 cohort who came from high-quality schools (60.2%). This pattern is consistent with our previous discussion on changes in the behavior of top *preuniversitarios* in the second cohort.

We note that while students who rejected the scholarship offer differ from students who accepted the offer in a number of characteristics, they are also quite similar in a number of socioeconomic and individual characteristics. This suggests that the decision to reject the scholarship is related more to academic outcomes than to socioeconomic characteristics or individual differences. In fact, as shown in Panel A, Table 7, students who rejected scholarship offers came on average from schools with higher SIMCE results in language and math, had a higher GPA, and performed better in the baseline tests (baseline language and math PSU score), all these differences being significant at the 1 per cent level.

In contrast, the two groups are not different in terms of gender, mother's education, indices of books and assets at home, and non-cognitive outcomes. Differences between students who accepted the scholarship and those who rejected it are particularly relevant for students in the second cohort. While we find few statistically significant differences between students who accepted or rejected the scholarship in the first cohort (Panel B in Table 7), we observe statistically significant differences in the quality of the schools attended and the academic performance of students who accepted or rejected the scholarship offer in 2011 (Panel C in Table 7).

---

[21] This variable equals 1 if the students declare in the follow-up survey that they have attended one of the three *preuniversitarios* we define as being of the highest quality (Preu UC, Cepech, and Pedro de Valdivia).

Regarding completion rates, the scholarship increased the likelihood of completing Preu UC by 65.6 p.p. (Table 11). In this case, we observe no significant differences between cohorts, suggesting that while students in the second cohort were less likely to accept the scholarship, once they did, they had a higher probability of completing the program. Overall, this implies that the impact of offering a scholarship over completion rates is non-statistically different between cohorts.

This, as previously discussed, could suggest that the targeting of scholarships at a population that could complete the program was much better done over the second year. Alternatively, it could be that in 2011, the *preuniversitario* was particularly relevant because it offered students who were affected by the student protests and lost school days a way of learning academic materials and preparing for the PSU test. In turn, with respect to heterogeneous effects in terms of school quality, we observe a similar pattern to that presented in column 3, with students from high-performing schools having a lower probability of completing the program.

This experiment builds on the fact that financial barriers may reduce the opportunities for low- and middle-income students to obtain high-quality preparation for PSU scores, which appears to be relevant in terms of access to higher education. While we discuss the latter in the following sections, here we address the former issue. Columns 9 to 12 in Table 11 present results for the impact of scholarship offers on attending a high-quality test preparation institution, where high-quality *preuniversitarios* are defined as the three *preuniversitarios* that are widely recognized as being of the highest quality among the available supply of massive *preuniversitarios* – including CEPECH, Pedro de Valdivia, and Preu UC.[22]

It should be noted that even if not offered a scholarship for Preu UC, students may still attend one of these institutions for their PSU preparation either by means provided by their households or by means of other scholarships, as previously discussed. In fact, data indicate that almost 50 per cent of the students not offered a scholarship by Preu UC still went to a top *preuniversitario*. In turn, the impact of the offer was 37.0 per cent (column 9).

Notably, differences in impact across subsamples are remarkable in this case. As shown in column 12, while students of the 2010 cohort who came from low-quality schools increased their access to top *preuniversitarios* by 51.1 per cent, students from the same cohort but from high-quality schools increased access by 33.7 per cent, students of the 2011 cohort and low-quality schools increased their access by 33.9 per cent, and students from that cohort but from high-quality schools increased their access by only 19 per cent. All the differences in impacts across subgroups are statistically significant at the 1 per cent level.

An alternative way of studying how the offer of a scholarship has heterogeneous effects on different types of students is to implement the methodology suggested by Kling (2001). We first estimate the probability of attending a top *preuniversitario*, only

[22] This variable equals 1 if the student declares in the follow-up survey that she attended one of the three preuniversitarios that are recognized as being of the highest quality.

including observations from the control group. The determinants of the probability of attending these institutions are the baseline PSU score, grades, family income, and a dummy indicating whether the student attends a high-quality school. Figure 5 presents estimates of the effect of the offer on the probability of attending a top *preuniversitario* for students located in different deciles of the distribution of the probability of attending a top *preuniversitario* for the complete sample and for each cohort.

Results confirm our previous discussion: the effect of receiving an offer is significantly higher among students who have, *ex ante*, a lower probability of attending a top *preuniversitario*, and decreases almost monotonically for students whose probability of attending a top *preuniversitario* is higher than 50 per cent. In fact, for students in the top decile, the offer of a scholarship does not change the probability of attending a top *preuniversitario*. The results for both cohorts maintain these patterns, but there is a decrease in the effect of the offer on students having a low probability of attending a top *preuniversitario* for the second cohort.

Results in Figure 5 are also useful in testing the implications of the monotonicity assumption needed to interpret the IV estimates as LATE effects. In the case of this study, the monotonicity assumption implies that receiving an offer should not decrease the probability of attending a top *preuniversitario* for any individual. This assumption has the testable implication that the difference in the probability of attending a top *preuniversitario* for the two values of a binary instrument should not be negative, which is satisfied in Figure 5, because we do not see a negative impact of the offer of a scholarship in any group of the population. In addition, this figure also suggests who are the most likely compliers and, therefore, the groups that will be driving the IV estimates (we will come back to this point later).

These results show that financial aid effectively increases access to high-quality PSU test preparation, and indeed, that financial restrictions operate as a barrier, especially for students from low-performing schools. The random assignment of a scholarship produces significant increases in takeup for most of the relevant groups, but it is much more important for subgroups that have a lower *ex ante* probability of attending a top *preuniversitario*. As previously discussed, our experiment is taking place in a context in which there is a private market operating and using several marketing tools to sell their services.

It is notable that in a context such as this, in which the service being offered is subject to several information asymmetries (i.e. clients cannot precisely determine the quality of the service they are purchasing), and in which the production of such service depends on the quality of the clients (i.e. the results from the *preuniversitarios* depend on the academic ability of their students), we can expect private firms to offer scholarships to high-performing students. The existence of a private market explains why the scholarship offer may have a lower impact on certain subgroups of students.

7.2 PSU preparation and results

We now move to estimating the impacts of the program on the process of PSU preparation in the 12th grade. This involves high school performance across 12th grade and PSU outcomes, including PSU participation and scores. With regard to the former, we start by estimating regressions for the impact of the program on grades,[23] and attendance and repetition rates. That our students attended 12th grade and *preuniversitario* during the whole year – most of them attending *preuniversitario* during the afternoon after school – implies that we could expect to see some substitution effects between one and the other, with students attending *preuniversitario* spending less time on school work. This factor is less likely to be binding over the second year of the experiment, when students missed a number of school days due to the protests.

Table 12 presents the results for these estimations. Column 1 in Panel A shows a small negative impact on students' grades, which is not statistically significant. However, as shown in column 2, the impact on grades is negative, statistically significant, and as high as 5.58 PSU points (0.08σ) for the 2010 cohort students who came from low-quality high schools. Instrumental variable results in columns 1 and 2 of Panel B show the same pattern, but with impacts of greater magnitude (9.33 PSU points, 0.14σ), showing that the stronger the participation of students in top *preuniversitarios*, the stronger the reduction in grades in the 2010 cohort.

Moreover, columns 3 and 4 of Panel A show that the pattern of impacts on the 12th grade attendance rate is similar to what is observed for grades. In fact, we estimate a negative average impact of 0.80 p.p. (0.06σ), which, as shown by column 4, is mostly concentrated in 2010 students who came from low-quality schools, for which the impact goes up to 1.52 p.p. (0.11σ). These two results jointly may be interpreted as follows: students in the first cohort somehow substituted between school and *preuniversitario*, as increased access to *preuniversitario* is associated with decreasing high school performance and attendance.

The natural question is, why does this not happen in the second cohort? Our hypothesis is that schools became less demanding in 2011 due to major student protests. While a number of students took over schools between June and October 2011, losing a considerable number of school days, students continued to attend Preu UC in the meantime. In fact, our data indicate that attendance rates for Preu UC in the period June–October are not statistically different between cohorts (in both years, attendance rates averaged 76%). We hypothesize that during 2011, the scholarship program offered students who were affected by the student protests a way of learning academic materials and preparing for the PSU test. The remedial effect of *preuniversitario* appears to have been particularly relevant for students attending low-quality schools, which is consistent with results that we present in what follows. Our hypothesis is that while the student protests were strong in high-quality schools, the protests had a higher impact in terms of learning loss among low-quality schools.

[23] Note that we transform grades, which in Chile work on a scale between 1 and 7 (with 4 being the passing grade) to a PSU-based scale, which is how grades are included in weighted scores that students use to apply to higher education institutions. This is to make results comparable with impacts on PSU scores.

To test this, we estimate the mapping between test scores at school level (using a test that took place before the student protests) and PSU scores in the 2010 and 2011 cohorts. We estimate panel data models using the following equation:

$$y_{st} = \kappa G_s * D_t + \eta_s + E_{st}$$

where y is average PSU score for school s at time t, Q_s is average SIMCE score (which does not vary by cohort),[24] D is a dummy that takes a value of one for the 2011 cohort, η are school fixed effects, and E is a shock. If κ is positive, it implies that the mapping between SIMCE scores (a proxy for school quality) and PSU scores is stronger for high-quality schools in 2011.

Table 13 presents evidence of this using information for all the public and voucher schools in Chile (the relevant comparison group for the subjects included in our experiment). Column 1 presents estimates, and the results imply that the mapping between SIMCE and PSU scores changes in 2011 and bad schools tend to decrease their performance relative to good schools. Column 2 presents a falsification exercise in which we report the same regression, but for private schools, which were not affected by the student protests. We do not find the same pattern that we found for public and voucher schools.[25] This evidence is consistent with our explanation for the different effects for students from low- and high-quality schools.

Next, we study the impacts of the scholarship program on students' rate of repetition of 12th grade, illustrated by columns 5 and 6 of Panel A. It is worth noting that repetition grades are quite low in Chile in normal years. However, in 2011 one of the forms of protest and support for the student protests was to voluntarily repeat 12th grade by rejecting the free (out-of-school) examinations the government offered by the government. While in 2010 the repetition rate in the control group of our sample was 0 per cent, the same variable increased to about 5.5 per cent in the second cohort. This increase in the repetition rates is almost entirely explained by an increase in the repetition rate from 0 per cent to about 10.6 per cent for students from high-quality schools. Instead, repetition rates remained 0 for students from low-quality schools. This probably reflects the fact that the students' movement was stronger in high-quality secondary schools in our sample.

Results show a statistically significant reduction of 4.0 p.p. (0.33σ) in the repetition rate of 2011 cohort students from high-quality schools. Note that in the first year of the intervention, there was no effect on repetition rates because, as already mentioned, repetition rates are close to 0 in Chile. That (i) the effect on repetition rates is more important in high-quality schools; (ii) we do not see large impacts on grades or attendance; and (iii) our sample is composed of above average students, implies that the effect of the treatment on repetition is probably related to a decrease in the

---

[24] We also tried a specification including yearly variations in Q, but due to collinearity across years in schools for the SIMCE scores, we chose to use only average Q for both cohorts.
[25] About 14 per cent of schools for which we have data on PSU and SIMCE results are private. The student protests did not affect these schools because students from private schools only participated in a couple of marches and did not lose school days as a consequence.

support of these students to the protests. This suggests that the opportunities open to students affect the political decisions they make. As discussed previously, the opinions of the students in the follow-up survey confirm this.

Next, we study the effect of the scholarships on the probability of taking the PSU exam. As previously discussed, in normal years, the coverage of the PSU exam among eligible students is close to 100 per cent, but takeup decreased in the second cohort, especially in low-performing schools. On the one hand we could expect to find a significant effect of the scholarship on PSU takeup rates for students from low-performing schools, because we believe that this group in particular benefited from the treatment. On the other hand, we could expect to find an impact on students from high-performing schools because the treatment helped to decrease repetition rates among this group.

Table 14 shows that this is not the case. Columns 1 and 2 suggest that the effects of the scholarship are negligible for the complete population and for all the subgroups included. It is important to emphasize that students can take the exam in one year and save their results for the application process of the next period. Then, a behavior in which a student takes the exam, chooses to repeat 12th grade as a way of supporting the student protests, and applies for higher education in the following year is possible. In all, these results suggest that the scholarship did not improve the takeup of the PSU exam in a significant way. The results also confirm our interpretation that the effect of the scholarships on repetition grades for students from high-quality schools corresponds more to a political choice rather than an effect of the scholarship on their academic outcomes.

In the remainder of Table 14, we study whether the scholarship offers affected the subject PSU tests that students took. As previously discussed, students can take science and history, with preparation for the former being more expensive, but also the area that allows students to apply to several of the most profitable tracks and universities (medicine, engineering, management, and economics). Thus, a possible effect of the scholarship offer is to change the subjects that students opt for.

Columns 3 and 4 show that we do not see an impact on the probability of taking an area (history or science) exam. In turn, results in columns 5 and 6 suggest that there seems to be a small decrease in the probability of taking the history exam (the probability of not taking the history exam increases). Together with the results in columns 3 and 4, this would imply that students are more likely to take the science exam. Note also that this effect is concentrated only in the group of students from low-quality schools; this is consistent with these being schools where students receive relatively weaker instruction levels in science.[26] However, results are not precisely estimated and, therefore, do not provide definitive evidence on this point.

[26] According to data from the SIMCE test scores in 2012, the difference in performance between students attending low and high socioeconomic status schools is higher in math (66 points) and science (63 points) than in language (50 points).

Next, we move on to studying our impact estimations on PSU results. Table 15 presents the results of such regressions. We start with language, an area for which results imply that the scholarship offers had no statistically significant average impact, as shown by column 1 of Panel A. In turn, column 2 shows a significant impact of 12.66 PSU points (0.17σ) on the 2011 cohort students who came from low-quality schools (C2, LQ). There is a similar pattern in the case of math. As shown in columns 3 and 4 of Panel A, there is a significant but small average impact of 3.91 PSU points (0.05σ) – significant at the 10 per cent level – which is, again, concentrated among students from low-quality schools, especially in the second cohort, showing impacts of 13.02 points (0.13σ) (C2, LQ).

Regarding the elective tests for science and history, we estimated impacts on these, correcting for self-selection in taking the tests using a Heckit procedure (Heckman 1979).[27] Given that we can only observe history and science scores for those students who took the science or history test, and considering that the program itself had an impact on the probability of taking the history or science test (see Table 15), we use a Heckit model to correct for selection bias. To do so, we first estimate a probit regression for the probability of taking the history or science test.[28] Next, we correct for self-selection by incorporating the inverse of the Mills ratio – a transformation of the predicted individual probabilities as an additional explanatory variable.

Columns 5 and 6 of Panel A show the results for science. We do not find a statistically significant average impact for the whole sample, but a marginally (p-value of 0.12) significant impact of 8.76 PSU points (0.11σ) for 2010 cohort students from low-quality schools. Similarly, columns 7 and 8 present results for the impact on history, for which, again, there is no significant average impact from the program, but a sizable impact of 41.50 PSU points (0.46σ) for 2011 students from low-quality schools.[29] Panel B in Table 15 shows the results from our IV estimations for the impact of attending a top *preuniversitario* on PSU scores. The pattern of results is similar to what is found for the impacts of offering scholarships, but the magnitude, again, is remarkably greater.[30]

---

[27] Results are almost unchanged if we run regressions without controlling *[contd. on p.31]* for potential sample selection bias.

[28] In this version, we do not use a specific excluded variable in the first stage of the Heckit procedure. However, in practice, some of the variables included as controls in the second stage are statistically significant in the first stage and not in the second stage, thus allowing us to identify the effect by not just using functional form assumptions.

[29] To save space, we do not report the inverse of Mill's ratio in the tables. The estimated coefficient is positive – and significant – for science and negative for history. As we expected, this implies that there is positive selection in science and negative selection in history.

[30] We also estimated the program impact using the number of correct answers as dependent variables in each case. Given that the PSU score is a nonlinear function of correct answers, there might have been differences between results in scores and in the number of correct answers. Appendix Table A2 presents the results and mostly confirms our estimated effects for scores, with the effects being more precisely estimated for math and less precisely for history.

In all, our results suggest that the access the scholarships provided, and the improved PSU preparation boosted by the program, impacted PSU performance. However, it did so particularly for students from low-quality schools, and more strongly for students in the 2011 cohort, who presented a bigger takeup rate in our experiment.

How do we interpret these results? First, the differences could be a consequence of changes in how the program was implemented. However, evidence that we could gather from classroom observations (Panel B of Table 10) suggests that this is not the case, it being hard to argue that program implementation significantly improved between years. Another possibility is that the treatment effects are different in the second cohort because the students in cohort 2011 differed in terms of socioeconomic and other characteristics from students in cohort 2010. Students in 2011 attended better high schools, had higher expectations in terms of PSU scores, and had slightly better high school performance and attendance rates on baseline (see Table 6, Panel A).

Considering that regressions control for all these characteristics, previous differences are only relevant if there are heterogeneous treatment effects across students who differ in terms of these characteristics. To check this, we ran a number of exercises in which we interacted the treatment effects with the variables that were unbalanced across cohorts. Appendix Table A1 presents the results of this exercise and suggests that differential interaction effects are not likely to explain the results, because only 6 out of 28 interaction effects are statistically significant. Moreover, in none of these cases do we find that the cohort effects change in a significant way. Therefore, differences in students' characteristics cannot explain different treatment effects across cohorts.

This leaves us with our final explanation. It is important to note that PSU scores provide an ordinal classification of students. Our results imply modest and only marginally significant improvements in 2010 in math and no effects on the other areas, and, at the same time, big and significant effects for students from the 2011 cohort who came from low-quality schools. To understand these changes, it is important to note that over the second year, student protests had an important effect on PSU scores, and this affected low-quality schools disproportionately. As discussed with respect to Table 13, student protests and the associated loss of school days and instruction seem to have affected low-performing schools more intensely.

Thus, our interpretation is that the scholarship program offered students who were the most affected by the student protests a way of learning academic materials and preparing for the PSU test. Putting it differently, these students faced competitors – regarding higher education access– who lost a significant number of days of their final school year and were not offered access to high-quality PSU preparation. Obviously, the external validity of these results is limited, because student protests are not common in many places and they do not take place every year. However, a more general interpretation of these results is that this type of scholarship may be a good way of helping students who attend schools that do not function properly. The rest of the students do not seem to benefit from this type of program.

7.3 Access to higher education

Previous results show that the scholarships for PSU preparation in effect translated into improved preparation and better PSU results for students in the 2011 cohort from low-quality high schools. Now we move toward estimating the impact of the scholarship on access to higher education. We start by estimating the impact on entering a higher education institution, and then estimate the impact on the type of institution students enter, and two proxies for the quality of the careers and institutions chosen by students in our sample.

Table 16 presents the results of our impact estimations on access to higher education. Column 1 of Panel A shows that students who were offered scholarships to Preu UC increased their access to higher education by 7.57 p.p. (0.17σ) from a mean of 72 per cent in the control group. This impact is mostly concentrated among students from the 2011 cohort, as shown in column 2. While students from the treatment group from low-quality schools increased their access to higher education by 14.61 p.p. (0.32σ), those from high-quality schools did so by 11.63 p.p. (0.25σ).

Moreover, IV estimations for the impact of attending a top *preuniversitario* show even larger results, as seen in Panel B in the same table. As displayed in column 1, the local average treatment effect of attending a top preparation institution on access to higher education was 15.4 p.p. (0.34σ), which, when estimated for cohort 2011 students, goes up to 42.44 and 42.92 p.p. (0.94σ and 0.95σ) for students from low- and high-quality schools, respectively. Indeed, results show that barriers to PSU preparation in effect reduce access to higher education. Moreover, access to high-quality preparation strongly reinforces such effects.

In contrast to the results on PSU outcomes, in this case we find impacts for students from both low- and high-quality schools. This is the result of impacts on PSU outcomes for low-quality students and a decrease in repetition rates for high-quality students. Also note that small changes in PSU outcomes may significantly affect entry to different programs, as Hastings, Neilson, and Zimmerman (2013) document.

In addition to the impacts on access to higher education, we also find that students in the treatment group were impacted in terms of the type of institutions at which they study. As previously discussed, three different types of higher education institutions exist in Chile: universities that offer standard undergraduate programs lasting about five years, IPs offering vocational degrees lasting four years, and CFTs offering vocational degrees lasting two years. In fact, as shown in columns 3 through 8 of Table 16, students' access to universities increases, which is partially compensated by a decrease in access to IPs. No impact is observed in terms of access to CFTs.

On average, students offered a scholarship for Preu UC increased their access to universities by 4.69 p.p. (0.10σ) from a mean of 66 per cent in the control group, and decreased their access to IPs by p.p. (0.09σ) from a mean of 17% in the control group. This result implies that these students, while still pursuing professional degrees in higher education, increased the quality of the institutions where they decided to study by switching from IPs to universities.

The pattern of impacts is the same as that found in terms of access to higher education, with almost all the impacts being focused among 2011 students, from both low- and high-quality schools. Magnitudes, again, are remarkable, with the difference in access to universities reaching 10.11 and 10.25 p.p. (0.21σ and 0.22σ) in the case of students from low- and high-quality schools, respectively; and the negative impact in access to IPs reaching 8.66 and 11.53 p.p. (0.22σ and 0.30σ), respectively.

Results for the impact of attending a top *preuniversitario* are qualitatively similar but, again, larger in magnitude. Thus, on average, students who attended a top *preuniversitario* because they received the scholarship offer increased their access to universities by 9.73 p.p. (0.20σ) more than those who did not attend, and, inversely, entered an IP by 8.51 p.p. (0.23σ) less than the latter. Impacts are, again, concentrated among 2011 students, and almost balanced between students from low- and high-quality high schools.

Indeed, 2011 cohort students from low-quality schools who attended a top *preuniversitario* increased their access to universities by 32.52 p.p. (0.69σ), while those from high-quality schools similarly increased access by 36.35 p.p. (0.76σ). On the other hand, the former decreased their access to IPs by 28.02 p.p. (0.77σ), and the latter did so by 39.36 p.p. (1.03σ). Thus, not only did treated students from the 2011 cohort increase their overall access to higher education, as mentioned earlier, they also switched from IPs to universities, which we interpret as a switch towards higher education institutions of higher quality.

Finally, in Table 17 we present regressions for two proxies for the quality of higher education tracks that students accessed. We use two indices that try to proxy for college quality (Hoxby 2009): one uses average PSU scores for students enrolled in each program, which is a combination of university and specific careers, and average expected wages of graduates from the program after they enter the job market. In both cases, we construct an index using the following procedure:

1. We collected information on average PSU scores and wages and employment rates for all the programs available at My Future. We collected information for 68 per cent of the students in the programs. To compute the wages of students who did not attend higher education, we use the expected wages after five years in the labor market for graduates from high school from MINEDUC (2013).

2. To estimate the weights of each index, we run regressions for members of the control group of each outcome on: (i) the results for each area of the PSU score (imputing a 0 to scores for students who did not take a subject exam); and (ii) dummies for students not taking a subject exam.

Appendix Table A3 presents the results of regressions of step 2. The results imply that average PSU in the program is increasing in all the PSU subject scores, with the higher weights given to math and language scores and to high school grades (which is a mechanical result of these being the portions of the tests with higher weights in the admissions formula in most programs).

In terms of area exams, both the weight for the science exam and the dummy for students not taking the history test (and, therefore, taking the science exam) have higher values than the weight given to the history exam and the dummy for not taking the history portion, respectively. This confirms that taking the science portions of the exam is a prerequisite for attending the more selective programs, because students who take and perform well in this exam increase their probability of attending a more selective program. The results for the expected (log) wages are slightly noisier, but also imply that the math scores receive a higher weight, followed by the language score.[31]

Table 17 presents the results for the treatment effects on our two measures of program quality. Columns 1 and 2 present results for average PSU scores. The results in column 1 of Panel A imply that students who were offered scholarships for Preu UC entered programs with about 2 PSU scores higher (0.03σ) than the control group, which is statistically and economically insignificant. However, this average impact masks a significant degree of heterogeneity, as column 2 suggests.

Students attending low-quality schools in the second cohort entered programs with about 10.5 PSU points (0.15σ) more than the control group. Effects for all the other groups are smaller and not different from 0. Moreover, IV estimations for the impact of attending a top *preuniversitario* show larger results, as is evident from Panel B of the same table. As displayed in column 2, the local average impact for students from low-quality schools in the second cohort is 27 PSU points (0.37σ).

Next, columns 3 and 4 present the same estimates for our index of expected labor income. On average, students who were offered scholarships entered the program with expected incomes of about 2.13 log points (0.03σ) above the control group, which is statistically significant but economically insignificant. However, again, there is a high degree of heterogeneity. Column 4 implies that students attending low-quality schools in the 2011 cohort entered programs with incomes about 6.83 log points (0.09σ) above the control group. As in column 4, IV estimates for the impact of attending a top *preuniversitario* show larger results: the local average impact for students from low-quality schools in the second cohort is 17.45 log points (0.23σ). As before, ITT and IV effects for all the other groups are not statistically different from 0 and much smaller in magnitude.

In sum, the results in this section imply that the offer of scholarships affected outcomes only in the second cohort, and especially for the group of students from low-quality schools. These effects are related to the probability of attending higher education and especially to the quality of program attended. The results for all the other groups are not different from 0, which is consistent with our previous results. The only exception to this pattern is the increase in the probability of attending university education for students from high-quality schools in the second cohort, which, as discussed, is probably explained by the decrease in the probability of repeating the 12th grade.

[31] That the results for income are noisier than results for PSU scores is because the income data are much noisier than the PSU scores and subject to measurement problems.

7.4 Impacts on perceptions

The data collected from our follow-up survey allow us to estimate impacts from the program on perceptions related to education. In particular, we study how the intervention affected students' perceptions regarding how public funds should be spent between different educational policies.

The idea is the following: to understand students' views on educational policy we asked them to prioritize a number of potential ways of spending public funds on education.[32] The motivation behind such an exercise was to understand if students could actually modify their perceptions on how to improve education through increased interaction with *preuniversitarios*, other educational institutions – for instance, Preu UC is based at PUC-Chile – and with the higher education admissions system itself.

Results are shown in Table 18. The only two policies for which our results are statistically significant are a reform of the higher education admissions system, the PSU, and free universal higher education, which, notably, was the main demand behind the 2011 student protests. While column 1 of Panel A shows that treated students would on average assign 16.5 per cent (0.17σ) more funds to reforming the higher education admissions system than students in the control group (who would assign Ch$1,189 to this item), that would be compensated by their assigning 10.2 per cent (0.10σ) less to achieving a free universal higher education system than would students in the control group (who would assign Ch$2,479 to this item), as shown in column 7. Our reading of these results is that, having been involved in strong PSU preparation during the year prior to being surveyed, students acknowledged difficulties in the admissions system and prioritized them in terms of policy relevance.

As in other results, there are heterogeneous results in this dimension too. The effect of an increase in expenditures for test preparation varies by year and cohort, and tends to match the impacts of the program. Students from low-quality schools in the first cohort – who did not improve their outcomes in a significant way – tend to prefer a higher expenditure in PSU preparation and a significantly lower expenditure on free higher education. In contrast, students from low-quality schools in the second cohort – who improved their outcomes significantly – do not have as strong views on this. In turn, students from high-quality schools who did not benefit from the scholarship program in terms of outcome do not present a clear picture of the impact of the program.

---

[32] The actual question was: "Imagine that you are the Minister of Education, and you are in charge of assigning Ch$10 billion between the following policy areas. How would you assign the funds?" Alternatives were: (i) to reform the higher education admission *[contd. on p. 36]* system; (ii) to increase expenditure on higher education aid through scholarships and free transportation; (iii) to invest in oversight of higher education institutions to reduce the extent of potential moneymaking; (iv) to invest in free universal higher education; and (v) to increase the quality of the higher education system. Ch$10 billion is equivalent to about US$20 million.

These results tend to confirm our previous discussion about the impact of the program on repetition rates. It seems that students react in terms of their decisions and policy views when they receive opportunities. In all, these results could have policy implications and deserve more study in future research.

7.5 Interpreting IV estimates

As discussed previously, our IV estimates identify a LATE for students who, due to the scholarship program, increased their probability of attending a top *preuniversitario*. In this section, we decompose the IV weighting following the analysis of Kling (2001). Table 19 presents the decomposition of IV weighting by the probability of attending a top *preuniversitario* decile. Following Kling (2001), the weight received by each decile q when using IV is

$$\omega_{q|X} = \frac{w_q \lambda_{q|X} \Delta TP_{q|X}}{\sum_q w_q \lambda_{q|X} \Delta TP_{q|X}}$$

where q represents deciles and X is a vector of control variables (given that our IV comes from a randomization, we use these variables just to improve the precision of the estimates). The weights depend on three different objects, which we now discuss. We present estimates of each magnitude in each column.

Column 1 presents estimates of $w_q = P(Q)$.

This, by definition of deciles, is 10 per cent for each group in our sample.

Column 2 presents

$$\lambda_{q|X} = E\left[P(Z|X, Q)\left(1 - P(Z|X, Q)\right)|Q\right]$$

where Z is the IV we use.

Given that Z comes from a randomization, λ is roughly constant across deciles, as presented in column 2. Finally, column 3 presents the estimates of

$$\Delta TP_{q|X} = E\left[E(TP|Z = 1, X, Q) - E(TP|Z = 0, X, Q)|Q\right]$$

which show the key element of $\omega_{q|X}$ (the estimates are in Figure 5). Then, in column 4 we report the implicit weights in the estimates. The results imply that, as expected, the deciles with the highest weights are those with a lower probability of attending a top *preuniversitario*, with weights above 10 per cent and as high as 17 per cent. In contrast, the top deciles have smaller weights.

8 Results from the tracking experiment

We now present the main results from the tracking experiment. According to the literature, tracking students into separate classes by prior academic achievement could have a negative impact on low-achieving students and benefit high-achieving students, if students benefit (are harmed) from having high- (low-) achieving peers through a spillover effect. On the other hand, tracking could benefit all students if it allows teachers to more closely match instruction to students' needs. In what follows, we attempt to estimate the impact of tracking in an out-of-school environment.

We begin by estimating the impact of attending tracking classes instead of attending mixed classes. Our main objective is to test whether attending tracking classes has a positive impact on students' achievement. Next, we look within tracking classes to test whether students in the middle of the distribution gain something from having higher-achieving peers. To do so, we use an RDD and compare the outcomes of the lowest-scoring student assigned to the high-achievement section to the outcomes of the highest-scoring students assigned to the low-achievement section. Finally, we estimate the impact of the mean and standard deviation of peers' performance on students' results. We do so by using an IV approach that exploits the allocation of students to different classes.

We start by estimating equation (5.4) by subject area (language, math, and science). The results in Table 20 show that there are no statistically significant effects from tracking classes for language, math or science. In fact, columns 1, 3, and 5 show average impacts of −0.268, 5.204, and 5.713 PSU points (−0.003σ, 0.07σ, and 0.08σ) in each of these areas, respectively, all of them non-significant. Moreover, in columns 2, 4, and 6 we estimate the impact of tracking separately for students from each of the two cohorts in the sample.

The results are, again, mostly non-significant, but they show a pattern similar to the results from the scholarships experiment. In fact, it seems like students of the 2011 cohort are more strongly favored by being assigned to tracked classes than those of the 2010 cohort. There are two marginally significant (with a p-value of 0.07 and 0.11) results: i.e. the impact of tracking on math and science for 2011 students, which are remarkably strong in magnitude, reaching as much as 9.41 and 11.14 PSU points (0.15 and 0.16σ), respectively.

Next, we estimate the effects of being in a high- versus low-quality class for students near median academic performance. As previously discussed, there are arguments in favor of a positive effect: traditional linear-in-means peer effects, higher teaching standards (if teachers adapt their strategies to the average student of the class; Pop-Eleches and Urquiola [2013]), and motivational effects. In contrast, there are also arguments in favor of a negative effect: a mismatch between student's abilities and the teaching level, motivational effects, and adaptation problems (Pop-Eleches and Urquiola 2013). We use an RDD with optimal bandwidth (Imbens and Kalyanaraman 2012) to estimate the effects.[33] Table 21 and Figure 4 present the results for the three tests. We find no significant effects for marginal students. We also estimate the models for the two cohorts and, consistent with the results in Table 20, we find different effects for the second cohort, but the effects are very imprecisely estimated.

[33] We also estimate RDD effects using the Calonico, Cattaneo and Titiunik (2014) procedure. Results are similar.

Finally, we estimate the effect of the mean and the variance of the quality of the peers in each classroom on PSU outcomes using an IV estimation procedure. The average and standard deviation of the performance of students at baseline are treated as endogenous variables, and the random allocation to mixed classes, together with the dummy for being allocated to a high-quality class (a quasi-random allocation based on a discontinuity),[34] are used as IVs. The results are presented in Table 22.

In the three cases, the endogenous variables have the expected signs: a positive effect on peer test scores and a negative impact on the standard deviation of this variable. However, the effects are only statistically significant at the 10 per cent level for the impact of the average test scores in math classes. The estimated impact implies that moving from the low-quality to the high-quality class increases test scores by about 13.3 PSU points (0.17σ). The results in science tests are similar in magnitude to the estimates for the math test, but are imprecisely estimated. Those for language are much smaller in magnitude.

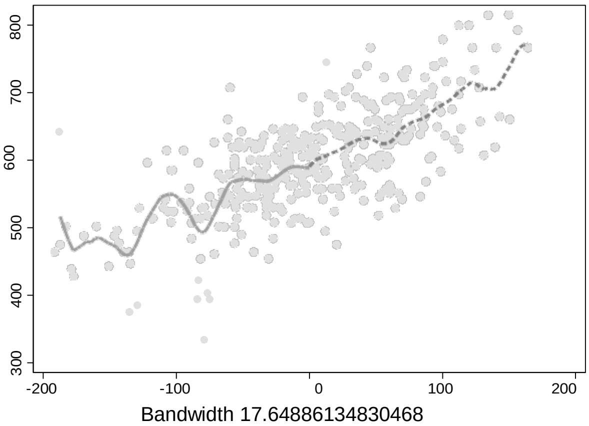
Overall, results on tracking prove to be null for language and a bit stronger for math and science, but low in magnitude and not statistically significant in most cases. Thus, these results do not support the findings of Duflo, Dupas and Kremer (2011) in Kenya. Even so, should we have expected to find similar results? A first argument against such a proposition would be that while their tracking intervention was implemented at the school level, where students shared the complete school day with their peers, our intervention involved only a couple of hours a week. Thus, the former was a much stronger intervention in regard to peers' interactions than the latter.

A second argument is built from the monitoring information we gathered through the implementation of the program. Duflo, Dupas and Kremer (2011) interpret their results as showing that reducing the variance of class composition allows teachers to adapt their teaching in a more appropriate way for students and thus boost their learning, literally teaching at their level. However, as previously discussed, we do not find systematic and big differences across different classes included in the tracking experiment (Panels D and E of Table 10).

In light of the mechanism that Duflo, Dupas and Kremer (2011) proposed for supporting the impact they found in their tracking experiment, these results imply that the absence of differences between tracked and mixed classes and also between high- and low-ability classrooms among the tracked classes in terms of teaching techniques and class dynamics might be a reason why we do not find any impact from tracking in our experiment. However, the small sample size in both sets of estimations limits the reach of this conclusion.

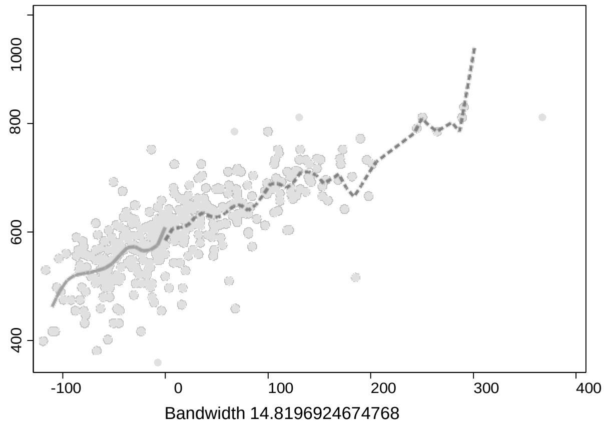
[34] We also control for cubic functions of the baseline of each test in each estimation.

Figure 4: RDD of the high-quality class impact

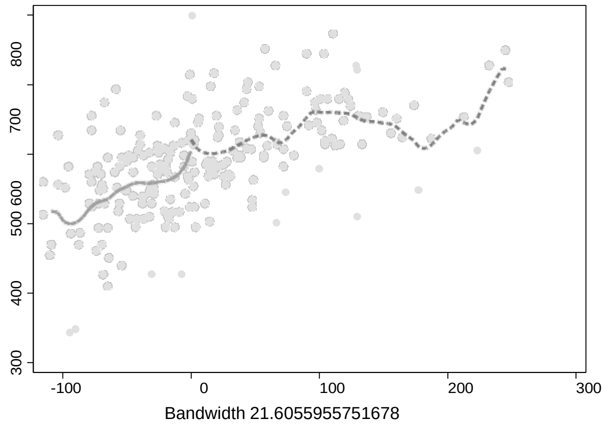


6.

(a) Language



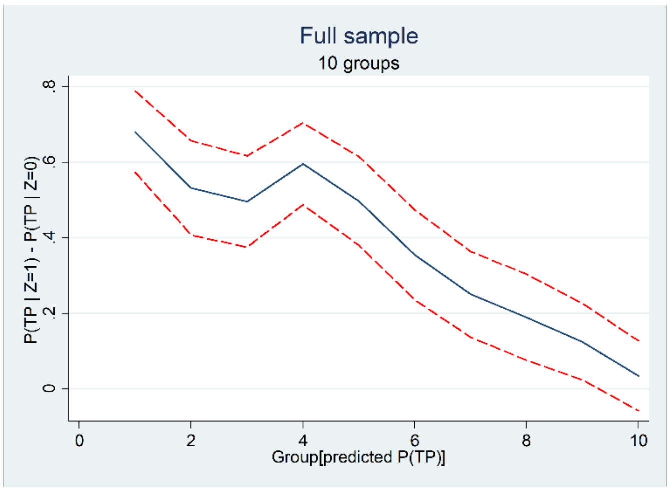
(b) Mathematics



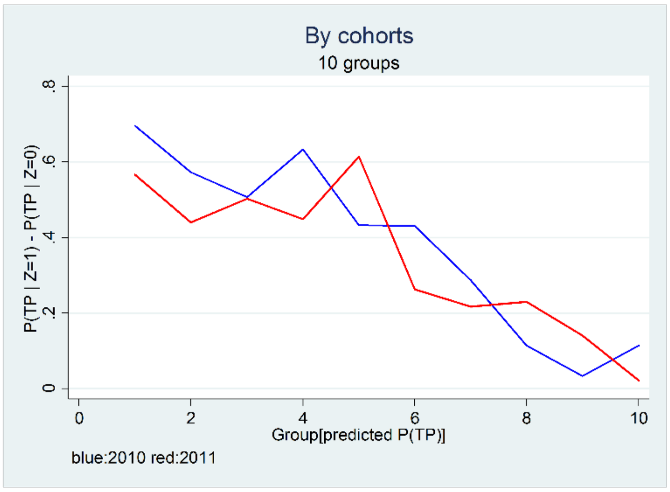
(c) Science

Note: This figure shows the tracking experiment impact using regression discontinuity design estimations. All the sample is used; see Table 21 Panel A for numeric estimates. Each dot represents one individual. The dependent variable is the student's PSU score, varying by subject. Graphs (a), (b), and (c) show language, math, and science scores, respectively. The independent variable is the student's PSU baseline score minus the baseline median score. The low- and high-ability class threshold is given by the median score. Kernel regressions using Imbens and Kalyanaraman (2012) optimal bandwidth are run at both sides of the discontinuity.

Figure 5: Effect of the offer on the probability of attending a top *preuniversitario* by subpopulations



(a) Full sample



(b) By cohort

Note: This figure presents estimates of the effect of the offer on the probability of attending a top *preuniversitario* for students located in different deciles of the distribution of the probability of attending a top *preuniversitario* for the complete sample and for each cohort. Subfigure (a) presents results for the full sample of students; subfigure (b) differentiates impact between cohorts. The dotted lines in figure (a) represent confidence intervals. For clarity, only the point estimates are graphed in subfigure (b).

9 Conclusions

Access to the most profitable tracks in higher education presents a steep socioeconomic gradient in most developed and developing countries. One potential explanation of this includes performance on standardized tests, an integral component of the college admissions process in many countries, in which wealthy students consistently outperform poor students, presumably because of their better educational background. This paper reports the results from a randomized experiment designed to evaluate the effects of offering scholarships for high-quality test preparation to high-achieving students on test outcomes and entry to higher education in Chile. The experiment took place in two different years, with the second year being affected by massive student protests.

We find that the scholarships significantly increased test preparation in high-quality institutions, especially among students from low-performing high schools. However, in terms of outcomes, while we find small and non-significant average impacts on test scores and higher education entry in the first cohort of students, we find a significant impact for treated students in the second cohort, especially among students from low-performing schools. Our interpretation is that over this period, in which students lost several school days, the scholarships provided a good option that substituted for preparation in school. This group of students also saw significant increases in the quality of higher education they accessed, entering into more selective tracks and with significantly higher expected labor market outcomes.

In addition, we find effects of the scholarships on the opinions and behaviors of students that relate to the policy discussion: students receiving scholarships supported higher public expenditures to improve test preparation for higher education and lower expenditures to decrease higher education fees. At the same time, students receiving scholarships in the second cohort supported the student protests less strongly. These results suggest that the opportunities that high school and university students receive affect their educational policy stance. This could have policy implications for several countries facing significant pressures from students to increase higher education expenditures.

Finally, we overlap a peer effects experiment in which scholarship students are allocated either to mixed-ability or tracking-by-ability classes. We do not find significant effects of this intervention on (blindly allocated) teacher and student outcomes.

Table 1: Entrance exam to tertiary education

Country	Exam
Australia	STAT, GAT
Brazil	Vestibular, ENEM
Chile	PSU
China	NCEE
Colombia	ICFES (Saber 11)
France	Baccalauréat Général
Germany	Abitur
Hungary	CEFR
India	AIEEE, CLAT, BITSAT, IIT-JEENEST, GGSIPU
Indonesia	SBMPTN, UMB, SBM-PTAIN
Israel	PET
Iran	Concours
Japan	University Center Test for University Admissions
Malaysia	SPM
Pakistan	NAT, GAT
Polonia	Matura
Russia	EGE
Saudi Arabia	Quadurat
Singapore	NCEE
Sweden	Högskoleprovet
South Korea	CSAT
Taiwan	NCEE
Turkey	YGS, LYS
United Kingdom	TSA
United States	SAT
Vietnam	Tuyen Sinh Dai Hoc - Cao Dang

Note: This table shows examples of entrance exams to college education varying by country.

Table 2: Annual expenditures on PSU exam preparation for people in the control group attending a top preuniversitario

	Median Expenditures	% of people who pay $0 US	% of people who pay less than $600 US	% of people who pay less than $1000 US	Obs
C1, LQ	900 US	5.80%	30.43%	57.97%	69
C1, HQ	785 US	14.44%	37.78%	67.78%	90
C2, LQ	640 US	12.50%	42.86%	78.57%	56
C2, HQ	736 US	10.47%	32.56%	66.28%	86

Note: This table shows expenditures on PSU preparation for students in the control group, by cohort and high school quality. Data are taken from the follow-up survey. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores.

Table 3: Sample composition of the scholarship experiment

Status	2010	2011	All
Applicants	1,116	961	2,077
Offered	632	652	1284
Accepted	488	420	908
Completed	430	412	842
Not Completed	58	8	66
Rejected	144	232	376
Not Offered	484	309	793

Note: Among the 376 students that rejected the offer, 197 attended a top preuniversitario - presumably because they received other (potentially better) scholarships, 144 did not attend a top preuniversitario, and 35 have missing data.

Table 4: Sample composition of the tracking experiment

		Subject		
Cohort	Class Level	Language	Math	Science
2010	High	87	92	41
	Low	86	92	40
	Mixed	89	82	39
2011	High	102	103	77
	Low	95	101	70
	Mixed	100	104	104
Total	High	189	195	118
	Low	181	193	110
	Mixed	189	186	143

Table 5: Sample vulnerability

School SES	Chile	RM	Experiment
Low	16.7	11.9	1.0
Middle-Low	31.0	28.8	15.1
Middle	33.0	35.9	43.1
Middle-High	18.6	22.6	39.7
High	0.8	0.7	1.1

Note: Data are at students' level and constructed using administrative data of MINEDUC in 2010. Only students in secondary grades, attending a non-private school and following an academic (non-vocational) track are considered.

Table 6: Balance between groups in the scholarship experiment

	Panel A - by Cohort		Panel B - by Treatment	
Variable	Mean C1	C2-C1	Mean C	T-C
Gender (= 1 if Female)	0.618	0.026	0.609	0.035
	(0.486)	(0.021)	(0.488)	(0.021)
School Dependence (= 1 if Public)	0.327	0.023	0.337	0.001
	(0.469)	(0.02)	(0.473)	(0.021)
School SIMCE, Language	279.096	5.205***	280.609	1.457
	(28.291)	(1.242)	(28.369)	(1.28)
School SIMCE, Math	281.925	7.894***	284.538	1.695
	(42.39)	(1.82)	(41.567)	(1.877)
Attendance Rate, 11th Grade	0.928	0.007***	0.932	0
	(0.062)	(0.002)	(0.061)	(0.002)
GPA, 11th Grade	5.971	0.082***	6.012	-0.003
	(0.325)	(0.014)	(0.331)	(0.014)
Baseline PSU Score, Language	505.011	41.04***	520.51	5.644*
	(65.65)	(2.956)	(67.056)	(3.169)
Baseline PSU Score, Math	451.362	-4.364	450.37	-1.661
	(85.146)	(3.568)	(83.26)	(3.663)
PSU Score Expectation, Language	647.622	7.361**	653.247	-3.598
	(80.269)	(3.48)	(78.927)	(3.574)
PSU Score Expectation, Math	644.471	13.807***	653.572	-4.395
	(100.832)	(4.377)	(96.277)	(4.501)
Attended Preu Previous Year	0.148	0.052***	0.186	-0.022
	(0.355)	(0.016)	(0.389)	(0.017)
Intends to Attend Preu	0.904	-0.019	0.899	-0.006
	(0.294)	(0.013)	(0.301)	(0.013)
Self-Control Index	3.555	0.057**	3.576	0.008
	(0.515)	(0.022)	(0.505)	(0.022)
Propensity to Plan	2.324	0.055	2.333	0.026
	(1.063)	(0.046)	(1.069)	(0.047)
Years of Education, Mother	12.679	0.307**	12.747	0.121
	(3.005)	(0.126)	(2.884)	(0.129)
Assets at Home Index	0.758	0.031***	0.777	-0.006
	(0.193)	(0.007)	(0.177)	(0.008)
Books at Home Index	4.258	0.205***	4.337	0.027
	(1.393)	(0.058)	(1.354)	(0.06)

Note: This table shows mean differences in baseline characteristics given by several assignments in the scholarship experiment. Standard errors are in parentheses. Panels A and B compare students by cohort and scholarship assignment, respectively. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. C and T denote students of the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 7: Balance by acceptance in the scholarship experiment

	Panel A - All Cohort 2011		Panel B - Cohort 2010		Panel C -	
Variable	Mean NR	R-NR	Mean NR	R-NR	Mean NR	R-NR
Gender (= 1 if Female)	0.643	0.00300	0.617	0.0710	0.674	-0.0530
	(0.479)	(0.029)	(0.487)	(0.046)	(0.469)	(0.039)
School Dependence (= 1 if Public)	0.307	0.110***	0.303	0.086*	0.312	0.123***
	(0.462)	(0.029)	(0.460)	(0.044)	(0.464)	(0.039)
School SIMCE, Language	280.5	5.377***	278.8	1.281	282.4	6.972***
	(27.879)	(1.732)	(28.221)	(2.724)	(27.384)	(2.254)
School SIMCE, Math	284.2	7.033***	281.7	0.439	287.1	9.712***
	(40.806)	(2.542)	(42.185)	(4.092)	(38.999)	(3.223)
Attendance Rate, 11th Grade	0.929	0.008**	0.927	0.00600	0.931	0.009**
	(0.057)	(0.003)	(0.059)	(0.005)	(0.055)	(0.004)
GPA, 11th Grade	5.986	0.078***	5.957	0.0480	6.019	0.080***
	(0.316)	(0.020)	(0.311)	(0.031)	(0.318)	(0.027)
Baseline PSU Score, Language	522.0	14.026***	503.6	1.339	543.5	11.902**
	(70.733)	(4.403)	(67.356)	(6.397)	(68.539)	(5.698)
Baseline PSU Score, Math	442.8	20.344***	448.0	10.92	436.6	29.038***
	(74.959)	(4.861)	(80.758)	(7.876)	(67.180)	(6.152)
PSU Score Expectation, Language	648.4	4.226	643.0	2.111	654.7	2.605
	(79.213)	(4.859)	(81.172)	(7.762)	(76.475)	(6.236)
PSU Score Expectation, Math	645.9	11.049*	639.7	5.594	653.3	11.04
	(99.977)	(6.229)	(103.954)	(9.837)	(94.722)	(8.095)
Attended Preu Previous Year	0.154	0.0360	0.139	0.0200	0.171	0.0370
	(0.361)	(0.023)	(0.347)	(0.033)	(0.377)	(0.032)
Intends to Attend Preu	0.884	0.0280	0.887	0.0150	0.881	0.0370
	(0.320)	(0.019)	(0.317)	(0.030)	(0.324)	(0.025)
Self-Control Index	3.576	0.0320	3.529	0.0310	3.631	0.00700
	(0.512)	(0.031)	(0.516)	(0.050)	(0.503)	(0.040)
Propensity to Plan	2.346	0.0490	2.338	-0.0160	2.356	0.0860
	(1.046)	(0.065)	(1.057)	(0.100)	(1.035)	(0.086)
Years of Education, Mother	12.79	0.272	12.70	0.0300	12.89	0.375*
	(2.869)	(0.175)	(3.066)	(0.293)	(2.624)	(0.214)
Assets at Home Index	0.767	0.0160	0.752	0.0120	0.784	0.0110
	(0.183)	(0.011)	(0.197)	(0.019)	(0.164)	(0.014)
Books at Home Index	4.371	-0.0240	4.330	-0.254*	4.419	0.0980
	(1.318)	(0.082)	(1.373)	(0.132)	(1.251)	(0.104)

Note: This table shows mean differences in baseline characteristics given by several assignments in the scholarship experiment. Standard errors are in parentheses. Panels A, B, and C compare students by scholarship acceptance. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. NR and R represent students that rejected and did not reject the scholarship offer. C and T denote students of the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 8: Balance between groups in the tracking experiment

	Panel A - Language		Panel B - Math		Panel C - Science	
Variable	Mean NT	T-NT	Mean NT	T-NT	Mean NT	T-NT
Gender (= 1 if Female)	0.582	0.107**	0.612	0.031	0.65	-0.053
	(0.494)	(0.042)	(0.488)	(0.043)	(0.478)	(0.051)
School Dependence (= 1 if Public)	0.312	0.031	0.306	0.018	0.321	-0.001
	(0.464)	(0.042)	(0.462)	(0.041)	(0.468)	(0.049)
School SIMCE, Language	281.671	-0.155	281.897	-2.825	283.387	-1.391
	(27.993)	(2.491)	(27.966)	(2.55)	(25.476)	(2.851)
School SIMCE, Math	286.1	-0.824	284.956	-2.162	289.267	-2.364
	(41.035)	(3.658)	(40.872)	(3.726)	(37.228)	(4.139)
Attendance Rate, 11th Grade	0.937	-0.007*	0.928	0.002	0.933	0.002
	(0.043)	(0.004)	(0.053)	(0.004)	(0.062)	(0.005)
GPA, 11th Grade	5.999	-0.009	5.986	0.017	6.052	-0.038
	(0.313)	(0.028)	(0.319)	(0.028)	(0.323)	(0.035)
Baseline PSU Score, Language	526.154	0.28	523.801	4.346	528.426	-0.136
	(70.135)	(6.285)	(68.715)	(6.302)	(64.739)	(7.353)
Baseline PSU Score, Math	443.39	-0.073	442.338	0.68	445.717	5.44
	(69.878)	(6.478)	(69.698)	(6.609)	(66.275)	(7.671)
PSU Score Expectation, Language	648.748	-0.564	659.924	-14.523**	642.965	1.911
	(80.259)	(7.137)	(77.269)	(7.01)	(84.63)	(8.456)
PSU Score Expectation, Math	650.686	-4.929	656.091	-15.383*	662.405	2.026
	(97.076)	(8.864)	(90.925)	(8.733)	(100.092)	(10.072)
Attended Preu Previous Year	0.18	-0.042	0.189	-0.052	0.118	0.053
	(0.385)	(0.032)	(0.392)	(0.032)	(0.324)	(0.038)
Intends to Attend Preu	0.899	0	0.897	-0.011	0.902	-0.029
	(0.301)	(0.026)	(0.303)	(0.027)	(0.298)	(0.034)
Self-Control Index	3.527	0.044	3.568	-0.01	3.615	-0.036
	(0.51)	(0.045)	(0.505)	(0.045)	(0.498)	(0.054)
Propensity to Plan	2.322	0.071	2.405	-0.001	2.258	0.04
	(1.008)	(0.092)	(1.084)	(0.095)	(0.954)	(0.108)
Years of Education, Mother	12.846	-0.002	12.853	-0.068	12.888	0.178
	(2.744)	(0.245)	(2.68)	(0.251)	(2.616)	(0.277)
Assets at Home Index	0.779	-0.014	0.767	-0.006	0.773	0.011
	(0.165)	(0.016)	(0.169)	(0.016)	(0.155)	(0.018)
Books at Home Index	4.455	-0.133	4.532	-0.241**	4.398	0.013
	(1.273)	(0.116)	(1.307)	(0.114)	(1.267)	(0.134)

Note: This table shows mean differences in baseline characteristics in the tracking experiment. Standard errors are in parentheses. Panels A, B, and C compare students by tracking assignment within language, math, and science classes, respectively. NT and T represent students in the non-tracking (mixed ability class) and tracking (high- or low-ability class) groups, respectively. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 9: Balance between attriters across survey and treatment status

	Panel A - by Survey Status		Panel B - Non Surveyed by Treatment Status	
Variable	S	NS-S	C	T-C
Gender (= 1 if Female)	0.663	-0.036	0.632	0.065
	(0.473)	(0.035)	(0.484)	(0.066)
School Dependence (= 1 if Public)	0.361	-0.024	0.32	0.085
	(0.481)	(0.035)	(0.468)	(0.067)
School SIMCE, Language	281.44	0.077	281.66	-0.462
	(29.143)	(2.097)	(30.015)	(4.116)
School SIMCE, Math	285.272	0.348	286.15	-1.848
	(42.498)	(3.074)	(43.633)	(6.001)
Attendance Rate, 11th Grade	0.928	0.003	0.93	-0.003
	(0.049)	(0.004)	(0.046)	(0.006)
GPA, 11th Grade	5.964	0.05**	5.992	-0.059
	(0.328)	(0.024)	(0.363)	(0.046)
Baseline PSU Score, Language	524.323	-0.358	528.764	-9.343
	(70.331)	(5.201)	(68.273)	(9.911)
Baseline PSU Score, Math	442.095	8.029	446.222	-8.685
	(77.467)	(6.004)	(79.749)	(10.924)
PSU Score Expectation, Language	653.638	-2.897	656.462	-5.941
	(82.094)	(5.854)	(78.884)	(11.587)
PSU Score Expectation, Math	642.47	9.29	646.462	-8.399
	(92.064)	(7.372)	(83.643)	(12.989)
Attended Preu Previous Year	0.163	0.01	0.207	-0.092*
	(0.37)	(0.028)	(0.407)	(0.051)
Intends to Attend Preu	0.836	0.064***	0.877	-0.085
	(0.37)	(0.022)	(0.329)	(0.051)
Self-Control Index	3.599	-0.019	3.61	-0.023
	(0.479)	(0.037)	(0.488)	(0.067)
Propensity to Plan	2.242	0.119	2.179	0.133
	(1.024)	(0.078)	(1.058)	(0.144)
Years of Education, Mother	12.826	-0.004	12.952	-0.265
	(2.804)	(0.212)	(2.879)	(0.395)
Assets at Home Index	0.729	0.048***	0.717	0.025
	(0.19)	(0.013)	(0.184)	(0.026)
Books at Home Index	4.153	0.222**	4.132	0.045
	(1.486)	(0.099)	(1.524)	(0.209)

Note: This table shows mean differences between attriters and non-attriters from the follow-up survey. Standard errors are in parentheses. NS and S represent surveyed and non-surveyed students. C and T denote students of the control and treatment groups. Panel A compares attriters and non-attriters. Panel B compares attriters by scholarship treatment assignment. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 10: Differences in classroom outcomes by cohort and experimental assignment

Variables	Mean of Scholarship Classes	Panel A Scholarship -Non-scholarship	Panel B Cohort 2 - Cohort 1	Panel C Track. Exp. - No Track. Exp.	Panel D Tracked - Mixed	Panel E High -Low
N° of Students in Class	22.60	2.356***	3.200***	1.806*	0.771	4.389***
	-8.225	(0.683)	(0.955)	(1.060)	(1.337)	(1.576)
Time for Teacher to Start Class (Mins)	3.714	-0.196	-0.732	-0.200	0.725	1.010
	(3.936)	(0.452)	(0.712)	(0.796)	(1.198)	(1.879)
Teacher's Control over Class (1–3 Scale)	1.595	0.101**	-0.0578	-0.0158	-0.0232	0.182
	(0.643)	(0.0501)	(0.0766)	(0.0854)	(0.107)	(0.123)
Punctual Teacher (= 1 if Punctual)	0.586	0.0308	-0.0980	-0.299***	-0.124	-0.0742
	(0.494)	(0.0576)	(0.0874)	(0.0934)	(0.128)	(0.178)
Late Students (= 1 if Yes)	0.761	-0.0743	-0.0366	-0.0254	0.0462	0.0281
	(0.428)	(0.0460)	(0.0756)	(0.0844)	(0.109)	(0.122)
Teacher Understands Contents (1–3 Scale)	2.377	0.0146	-0.0771	-0.143**	-0.117	-0.237**
	(0.519)	(0.0420)	(0.0618)	(0.0688)	(0.0895)	(0.106)
Teacher Integrates Students (= 1 if Yes)	0.414	0.0559	0.0325	-0.0510	-0.115	0.00589
	(0.494)	(0.0419)	(0.0630)	(0.0711)	(0.0918)	(0.112)
Teacher Adapts Content to Students' Needs (= 1 if Yes)	0.238	-0.0386	0.132***	-0.0747	0.0439	-0.162*
	(0.426)	(0.0355)	(0.0506)	(0.0563)	(0.0731)	(0.0888)
Teacher Talks about Recurrent Mistakes (= 1 if Yes)	0.257	0.00218	0.105**	-0.0543	0.0341	-0.263***
	(0.438)	(0.0355)	(0.0522)	(0.0582)	(0.0760)	(0.0888)
Students' Attitude in Class (1–4 Scale)	2.495	0.0471	0.220**	0.0861	0.0284	-0.0338
	(0.889)	(0.0719)	(0.106)	(0.119)	(0.143)	(0.187)
Teacher Asks Students (= 1 if Yes)	0.881	-0.00929	0.0511	-0.0517	-0.0730	-0.0856
	(0.324)	(0.0261)	(0.0379)	(0.0423)	(0.0549)	(0.0737)
Students Answer Questions (1–4 Scale)	1.758	0.109*	-0.0978	0.101	-0.170	0.00415
	(0.790)	(0.0630)	(0.0985)	(0.111)	(0.128)	(0.167)
Students Asks Questions (1–4 Scale)	2.974	0.0505	-0.166*	0.103	-0.387***	-0.288
	(0.848)	(0.0680)	(0.0995)	(0.111)	(0.142)	(0.175)
Teacher Answers Questions (1–4 Scale)	3.384	0.0652	-0.158**	-0.0947	0.0580	-0.211
	(0.543)	(0.0546)	(0.0760)	(0.0848)	(0.114)	(0.134)
Teacher Disrespectful with Students (= 1 if Yes)	0.0261	0.0167	0.0161	-0.0336	0.0218	-0.0381
	(0.160)	(0.0109)	(0.0191)	(0.0213)	(0.0203)	(0.0313)
Students Disrespectful with Teacher (= 1 if Yes)	0.0987	0.00576	-0.0579	0.00957	-0.00758	0.159**
	(0.299)	(0.0241)	(0.0357)	(0.0400)	(0.0545)	(0.0653)
Student–Teacher Relationship (1–4 Scale)	2.090	-0.000802	-0.175**	0.225***	-0.0172	-0.141
	(0.578)	(0.0482)	(0.0702)	(0.0787)	(0.0987)	(0.125)
Teacher Uses Projector (1–4 Scale)	1.414	0.0603	0.201*	0.0622	-0.330**	0.452***
	(0.997)	(0.0733)	(0.106)	(0.118)	(0.154)	(0.169)
Teacher Sends Homework (= 1 if Yes)	0.365	-0.0634	0.0189	-0.0479	-0.0698	-0.0143
	(0.482)	(0.0402)	(0.0587)	(0.0669)	(0.0864)	(0.106)
Teacher Provides Worksheets (= 1 if Yes)	0.941	-0.00145	0.0611**	0.0402	-0.0190	-0.00318
	(0.236)	(0.0186)	(0.0280)	(0.0311)	(0.0303)	(0.0403)
Teacher References Worksheets (= 1 if Yes)	0.712	-0.0387	-0.270***	-0.0694	0.0470	-0.0572
	-0.454	(0.0340)	(0.0524)	(0.0583)	(0.0777)	(0.0966)

Note: This table shows differences in classroom outcomes by several assignments. The unit of observation is a classroom. Reported coefficients come from OLS regressions. All regressions control for class subject. Regressions in Panels A, C, D, and E include cohort dummies. Standard errors are in parentheses. Panel A compares scholarship classes with other classes of Preu UC including only non-scholarship students who are not part of this evaluation. Panel B compares classes in cohorts 2010 and 2011. Panel C compares classes that were assigned to the tracking experiment with classes that were not assigned to the tracking experiment. Panel D compares tracking classes, including low- and high-ability classes, with mixed classes in the tracking experiment. Panel E compares high- and low-ability tracked classes. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 11: Impact estimations for the scholarship experiment on access to test preparation

Dep. Var.	Accepted Scholarship for Preu UC				Completed Preu UC				Attended a Top *Preuniversitario*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
T	0.707***	0.772***	0.752***	0.802***	0.656***	0.680***	0.687***	0.695***	0.370***	0.433***	0.452***	0.511***
	(0.016)	(0.021)	(0.023)	(0.029)	(0.017)	(0.023)	(0.024)	(0.031)	(0.019)	(0.025)	(0.026)	(0.034)
T x C2		-0.128***		-0.107**		-0.048		-0.017		-0.172***		-0.172***
		(0.033)		(0.046)		(0.034)		(0.048)		(0.038)		(0.054)
T x HQ			-0.086***	-0.060			-0.061*	-0.029			-0.173***	-0.174***
			(0.032)	(0.043)			(0.034)	(0.045)			(0.038)	(0.050)
T x C2 x HQ				-0.033				-0.056				0.025
				(0.065)				(0.069)				(0.076)
Mean of Control	0	0	0	0	0	0	0	0	.499	.499	.499	.499
Observations	2,077	2,077	2,072	2,072	2,077	2,077	2,072	2,072	1,939	1,939	1,934	1,934
R-squared	0.480	0.490	0.485	0.495	0.421	0.423	0.424	0.426	0.162	0.186	0.184	0.204
Impacts by Subsample												
C1		0.772***				0.68***				0.433***		
C2		0.644***				0.632***				0.261***		
LQ			0.752***				0.687***				0.452***	
HQ C1,			0.666***				0.626***				0.279***	
LQ				0.802***				0.695***				0.511***
C1, HQ				0.742***				0.666***				0.337***
C2, LQ				0.695***				0.678***				0.339***
C2, HQ				0.602***				0.593***				0.19***

Note: This table shows the impact of the scholarship assignment on access to PSU exam preparation. Reported coefficients come from OLS regressions. All regressions include cohort dummies and control for class subject. Standard errors are in parentheses. T is the estimate of the impact of offering a scholarship. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined as being above or below the median school in the sample in terms of average 10th grade SIMCE test scores. The impact by subsample shows the T estimate when the sample is restricted by cohorts, high school qualities, and interactions between the two. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 12: Impact estimations for the scholarship experiment on 12th grade performance

Dep. Var.	Grades		Attendance (3)		Repetition Rate	
	(1)	(2)		(4)	(5)	(6)
	Panel A - Impact of Offering a Scholarship for Preu UC					
T	-2.303	-5.584**	-0.00797	-0.0152*	-0.00989*	0.00650
	(1.614)	(2.775)	(0.00512)	(0.00892)	(0.00592)	(0.0103)
T x C2		6.922*		0.0151		-0.00875
		(4.189)		(0.0134)		(0.0154)
T x HQ		4.160		0.00534		-0.00783
		(4.009)		(0.0129)		(0.0149)
T x C2 x HQ		-7.315		-0.00791		-0.0400**
		(5.378)		(0.0171)		(0.0197)
Mean of Control	630	630	.884	.884	.0209	.0209
Observations	1,928	1,928	1,994	1,994	1,995	1,995
R-squared	0.805	0.806	0.573	0.573	0.336	0.342
Impacts by subsample						
C1, LQ		-5.584**		-0.0152*		0.0065
C1, HQ		-1.424		-0.0099		-0.0013
C2, LQ		1.338		-0.0001		-0.0023
C2, HQ		-1.817		-0.0027		-0.0501***
	Panel B - Impact of Attending a Top Preuniversitario					
TP	-5.597	-9.328*	-0.0111	-0.0185	-0.0341**	0.0283
	(4.483)	(5.568)	(0.0132)	(0.0164)	(0.0171)	(0.0218)
TP x C2		9.035		0.0231		-0.0834**
		(10.61)		(0.0311)		(0.0414)
TP x HQ		6.708		-0.00452		-0.0636
		(10.29)		(0.0301)		(0.0401)
TP x C2 x HQ		-11.64		0.0125		-0.0998***
		(9.170)		(0.0270)		(0.0359)
Mean of Control	630	630	.884	.884	.0209	.0209
Observations	1,808	1,808	1,864	1,864	1,865	1,865
Impacts by subsample						
C1, LQ		-9.328*		-0.0185		0.0283
C1, HQ		-2.62		-0.023		-0.0353
C2, LQ		-0.293		0.0046		-0.0551
C2, HQ		-5.225		0.0126		-0.2185***

Note: This table shows the scholarship experiment impact on student's grades, class attendance, and repetition rates. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT), and from IV regressions for Panel B (LATE). All regressions include dummies for cohort and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the T and TP estimate when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 13: Correlation between school's SIMCE and PSU results for each cohort

Dep. Var.	School's PSU Score	
	(1)	(2)
Cohort	-12.33***	4.251
	(3.881)	(16.18)
SIMCE x Cohort	0.0367**	-0.0260
	(0.0151)	(0.0514)
Constant	470.8***	603.4***
	(0.338)	(0.898)
Observations	4,273	697
R-squared	0.971	0.962

Note: Cohort is a dummy equal to 1 when the cohort is 2011. The unit of observation is a school. Reported coefficients come from OLS regressions. The dependent variable is an average between language and math PSU scores. Column 1 shows results for public and private subsidized schools; column 2 considers only private paid (non-subsidized) schools. Standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 14: impact estimations for the scholarship experiment on PSU participation

Dep. Var.	PSU Not Taken		PSU Elective Not Taken		PSU History Not Taken	
	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A - Impact of Offering a Scholarship for Preu UC					
T	-0.00317	0.00825	-0.00527	0.00875	0.0239	0.0646
	(0.00698)	(0.0122)	(0.00750)	(0.0131)	(0.0231)	(0.0401)
T x C2		-0.0232		-0.0222		0.0227
		(0.0182)		(0.0196)		(0.0601)
T x HQ		-0.00770		-0.0102		-0.0590
		(0.0176)		(0.0189)		(0.0581)
T x C2 x HQ		0.00946		0.00138		-0.0990
		(0.0233)		(0.0251)		(0.0770)
Mean of Control	.0209	.0209	.0261	.0261	.576	.576
Observations	1,995	1,995	1,995	1,995	1,995	1,995
R-squared	0.209	0.210	0.204	0.205	0.178	0.181
Impacts by Subsamples						
C1, LQ		0.00825		0.00875		0.0646
C1, HQ		0.00055		-0.00145		0.0056
C2, LQ		-0.01495		-0.01345		0.0873*
C2, HQ		-0.01319		-0.02227		-0.0707
	Panel B - Impact of Attending a Top Preuniversitario					
TP	-0.00184	0.00572	-0.00851	0.00895	0.0724	0.133*
	(0.0178)	(0.0220)	(0.0195)	(0.0241)	(0.0647)	(0.0801)
TP x C2		-0.0148		-0.0218		0.0467
		(0.0418)		(0.0457)		(0.152)
TP x HQ		-0.00640		-0.0212		-0.150
		(0.0405)		(0.0442)		(0.147)
TP x C2 x HQ		-0.00439		-0.0223		-0.165
		(0.0363)		(0.0396)		(0.132)
Mean of Control	.0209	.0209	.0261	.0261	.576	.576
Observations	1,865	1,865	1,865	1,865	1,865	1,865
Impacts by Subsamples						
C1, LQ		0.00572		0.00895		0.133*
C1, HQ		-0.00068		-0.01225		-0.0176
C2, LQ		-0.00908		-0.01285		0.1797
C2, HQ		-0.01987		-0.05635		-0.1353

Note: This table shows the scholarship experiment impact on PSU participation. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from IV regressions for Panel B (LATE estimates). All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 15: Impact estimations for the scholarship experiment on PSU score

Dep. Var.	Language PSU Score		Math PSU Score		Science PSU Score		History PSU Score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A - Impact of Offering a Scholarship for Preu UC								
T	1.909	2.773	3.912*	5.484	4.544	8.759	2.999	6.247
	(2.522)	(4.346)	(2.265)	(3.908)	(3.258)	(5.648)	(6.001)	(10.62)
T x C2		9.887		7.533		-1.855		35.25**
		(6.563)		(5.893)		(8.235)		(14.79)
T x HQ		-10.82*		-4.536		-4.449		-15.53
		(6.282)		(5.644)		(8.092)		(15.42)
T x C2 x HQ		1.000		-11.86		-6.691		-47.79**
		(8.395)		(7.540)		(10.71)		(20.74)
Mean of Control	594	594	602	602	587	587	594	594
Observations	1,946	1,946	1,946	1,946	1,235	1,235	773	773
R-squared	0.651	0.652	0.748	0.749	0.669	0.670	0.588	0.593
Impacts by Subsamples								
C1, LQ		2.773		5.484		8.759		6.247
C1, HQ		-8.047*		0.948		4.31		-9.283
C2, LQ		12.66**		13.017***		6.904		41.497***
C2, HQ		2.84		-3.379		-4.236		-21.823
Panel B - Impact of Attending a Top Preuniversitario								
TP	3.987	0.421	9.392	9.413	11.60	17.28	15.16	21.36
	(7.025)	(8.825)	(6.223)	(7.694)	(9.097)	(10.91)	(14.47)	(20.61)
TP x C2		42.68**		19.53		2.416		71.17**
		(16.78)		(14.66)		(21.41)		(33.51)
TP x HQ		-29.67*		-10.98		-11.64		-58.00
		(16.22)		(14.18)		(19.75)		(40.58)
TP x C2 x HQ		6.432		-18.12		-17.45		-70.85**
		(14.51)		(12.69)		(17.92)		(33.33)
Mean of Control	594	594	602	602	587	587	594	594
Observations	1,827	1,827	1,826	1,826	1,158	1,158	724	724
Impacts by Subsamples								
C1, LQ		0.421		9.413		17.28		21.36
C1, HQ		-29.249*		-1.567		5.64		-36.64
C2, LQ		43.101***		28.943**		19.696		92.53***
C2, HQ		19.863		-0.157		-9.394		-36.32

Note: This table shows the scholarship experiment impact on PSU score. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from instrumental variable regressions for Panel B (LATE estimates). All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. Results for science and history are estimated correcting for self-selection into the test. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 16: Impact estimations for the scholarship experiment on access to higher education

Dep. Var.	Enrolled in HE		Enrolled in University		Enrolled in IP		Enrolled in CFT	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A - Impact of Offering a Scholarship for Preu UC								
T	0.0757***	0.0489	0.0469**	0.0143	-0.0326**	0.0164	0.00642	0.0136
	(0.0203)	(0.0354)	(0.0211)	(0.0367)	(0.0162)	(0.0282)	(0.00563)	(0.00981)
T x C2		0.0972*		0.0868		-0.103**		-0.0120
		(0.0531)		(0.0550)		(0.0422)		(0.0147)
T x HQ		-0.0305		-0.0180		0.00441		-0.00764
		(0.0513)		(0.0532)		(0.0408)		(0.0142)
T x C2 x HQ		0.000662		0.0194		-0.0331		0.00688
		(0.0679)		(0.0704)		(0.0541)		(0.0188)
Mean of Control	.713	.713	.665	.665	.171	.171	.00913	.00913
Observations	1,995	1,995	1,995	1,995	1,995	1,995	1,995	1,995
R-squared	0.186	0.189	0.235	0.237	0.310	0.316	0.114	0.114
Impacts by Subsamples								
C1, LQ		0.0489		0.0143		0.0164		0.0136
C1, HQ		0.0184		-0.0037		0.0208		0.006
C2, LQ		0.1461***		0.1011**		-0.0866**		0.0016
C2, HQ		0.1163***		0.1025**		-0.1153***		0.0008
Panel B - Impact of Attending a Top Preuniversitario								
TP	0.154***	0.0524	0.0973*	0.0132	-0.0851*	0.0238	0.00864	0.0202
	(0.0565)	(0.0710)	(0.0585)	(0.0734)	(0.0455)	(0.0570)	(0.0151)	(0.0188)
TP x C2		0.372***		0.312**		-0.304***		-0.0293
		(0.135)		(0.139)		(0.108)		(0.0356)
TP x HQ		-0.0724		-0.0877		0.0106		-0.00613
		(0.130)		(0.135)		(0.105)		(0.0344)
TP x C2 x HQ		0.0772		0.126		-0.124		-0.000542
		(0.117)		(0.121)		(0.0939)		(0.0309)
Mean of Control	.713	.713	.665	.665	.171	.171	.00913	.00913
Observations	1,865	1,865	1,865	1,865	1,865	1,865	1,865	1,865
Impacts by Subsamples								
C1, LQ		0.0524		0.0132		0.0238		0.0202
C1, HQ		-0.02		-0.0745		0.0344		0.0141
C2, LQ		0.4244***		0.3252**		-0.2802***		-0.0091
C2, HQ		0.4292***		0.3635**		-0.3936***		-0.0158

Note: This table shows the scholarship experiment impact on access to higher education. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from IV regressions for Panel B (LATE estimates). All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 17: Impact estimations for the scholarship experiment on predicted quality of higher education institution

Dep. Var.	Predicted Average PSU Score		Predicted log(Wage)	
	(1)	(2)	(3)	(4)
Panel A - Impact of Offering a Scholarship for Preu UC				
T	2.038	2.003	0.0213*	0.0290
	(1.563)	(2.683)	(0.0120)	(0.0206)
T X C2		8.524**		0.0393
		(4.049)		(0.0311)
T X HQ		-3.493		-0.0341
		(3.874)		(0.0298)
T X C2 X HQ		-8.703*		-0.0370
		(5.195)		(0.0400)
Mean of Control	599	599	6.69	6.69
Observations	1,926	1,926	1,926	1,926
R-squared	0.830	0.831	0.881	0.881
Impacts by Subsamples				
C1, LQ		2.003		0.0290
C1, HQ		-1.49		-0.0051
C2, LQ		10.527***		0.0683***
C2, HQ		1.824		0.0313
Panel B - Impact of Attending a Top Preuniversitario				
TP	5.415	3.660	0.0488	0.0435
	(4.220)	(5.250)	(0.0324)	(0.0401)
TP X C2		23.34**		0.131*
		(10.04)		(0.0768)
TP X HQ		-10.94		-0.0788
		(9.727)		(0.0744)
TP X C2 X HQ		-13.12		-0.0612
		(8.679)		(0.0664)
Mean of Control	599	599	6.69	6.69
Observations	1,807	1,807	1,807	1,807
Impacts by Subsamples				
C1, LQ		3.660		0.0435
C1, HQ		-7.28		-0.0353
C2, LQ		27.00***		0.1745**
C2, HQ		13.88		0.1133

Note: This table shows the scholarship experiment impact on predicted quality of higher education institution. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from IV regressions for Panel B (LATE estimates). Dependent variable is college quality for which two different proxies are used, details of the construction in the text. All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 18: Impact estimations for the scholarship experiment on educational policy priorities

Dep. Var.: Expenditure in:	PSU Reform		HE Support		HE Oversight		Free HE		In HE Quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A- Impact of Offering a Scholarship for Preu UC										
T	199.7***	339.8***	-116.7	-8.371	-37.00	44.36	-252.9*	-352.1	29.25	9.327
	(70.05)	(111.9)	(113.9)	(190.4)	(89.04)	(144.5)	(150.8)	(241.1)	(125.5)	(209.8)
T X C2		-174.0		-73.56		-122.3		-266.6		-209.5
		(196.7)		(303.0)		(242.2)		(417.5)		(335.1)
T X HQ		-313.9*		-150.1		-197.6		237.7		187.7
		(163.7)		(276.8)		(211.0)		(352.0)		(305.0)
T X C2 X HQ		372.0		-45.71		298.8		442.6		40.68
		(235.7)		(380.2)		(300.2)		(519.3)		(419.2)
Mean of Control	1207	1207	2998	2998	1662	1662	2544	2544	2480	2480
Observations	1,543	1,543	1,687	1,687	1,583	1,583	1,524	1,524	1,679	1,679
R-squared	0.279	0.281	0.182	0.182	0.180	0.181	0.205	0.206	0.154	0.154
Impacts by Subsamples										
C1, LQ		399.8***		-8.371		44.36		-352.1		9.327
C1, HQ		25.9		-158.5		-153.2		-114.4		197.0
C2, LQ		165.8		-81.9		-77.9		-618.7*		-200.2
C2, HQ		537.8		-127.6		220.9		-176.1		-159.5
Panel B- Impact of Attending a Top Preuniversitario										
TP	536.2***	620.3***	-359.2	18.71	-104.1	69.95	-707.5**	-728.9*	-40.91	70.14
	(190.7)	(221.4)	(306.7)	(372.2)	(228.2)	(277.6)	(335.4)	(414.0)	(293.7)	(354.1)
TP X C2		241.0		-618.4		-150.0		-837.7		-865.5
		(609.3)		(778.1)		(596.6)		(820.5)		(758.6)
TP X HQ		-498.8		-562.1		-501.8		487.7		413.1
		(424.5)		(701.9)		(521.9)		(773.0)		(667.1)
TP X C2 X HQ		530.3		-221.8		374.1		560.1		-331.1
		(399.3)		(653.7)		(492.5)		(753.5)		(621.7)
Mean of Control	1207	1207	2998	2998	1662	1662	2544	2544	2480	2480
Observations	1,533	1,533	1,673	1,673	1,569	1,569	1,511	1,511	1,664	1,664
Impacts by Subsamples										
C1, LQ		620.3***		18.71		69.95		-728.9*		70.14
C1, HQ		121.5		-543.4		-431.85		-241.2		483.2
C2, LQ		861.3		-599.7		-80.05		-1566.6**		-795.4
C2, HQ		1391.6		-821.5		294.1		-1000.5		-1126.5

Note: This table shows the scholarship experiment impact on predicted quality of higher education institution. T is the estimate of the impact of offering a scholarship, and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from IV regressions for Panel B (LATE estimates). All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 19: Decomposition of IV weighting by population subgroup

	q (1)	$\lambda_{q\|x}$ (2)	$\Delta TP_{q\|x}$ (3)	$\omega_{q\|x}$ (4)
q = 1	0.10	0.24	0.68	0.19
q = 2	0.10	0.24	0.53	0.15
q = 3	0.10	0.23	0.50	0.13
q = 4	0.10	0.23	0.59	0.16
q = 5	0.10	0.23	0.50	0.13
q = 6	0.10	0.24	0.35	0.10
q = 7	0.10	0.23	0.25	0.06
q = 8	0.10	0.23	0.18	0.05
q = 9	0.10	0.23	0.13	0.03
q = 10	0.10	0.24	0.01	0.00

Note: $\omega_{q|x} = (w_q \lambda_{q|x} \Delta TP_{q|x}) / (\sum_q w_q \lambda_{q|x} \Delta TP_{q|x})$

Table 20: Impact estimations for the tracking experiment on PSU performance

Dep. Var.: PSU Score	Language (1)	Language (2)	Math (3)	Math (4)	Science (5)	Science (6)
Tracking	-0.268	-6.981	5.204	0.0900	5.713	-6.336
	(4.826)	(6.935)	(4.256)	(6.298)	(5.763)	(10.44)
Tracking X C2		12.94		9.321		17.48
		(9.602)		(8.462)		(12.63)
Observations	531	531	542	542	342	342
R-squared	0.580	0.582	0.673	0.674	0.556	0.559
Impacts by Subsamples						
C1		-6.981		0.0900		-6.336
C2		5.959		9.411*		11.144

Note: This table shows the tracking experiment impact on PSU performance. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. Reported coefficients come from OLS regressions. All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. The impact by subsample shows the tracking estimate when the sample is restricted by cohorts. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 21: Regression discontinuity design estimates of the high-quality class impact

Dep. Var.: PSU Score	Language (1)	Math (2)	Science (3)
All Sample			
High-quality class	1.919	-24.19	16.44
	(21.12)	(18.61)	(40.03)
Only Cohort 2010			
High-quality class	-6.517	-9.926	-29.16
	(33.80)	(25.04)	(45.03)
Only Cohort 2011			
High-quality class	9.594	2.750	22.80
	(22.50)	(26.86)	(49.28)

Note: This table shows the impact of being in a high-quality class for marginal students in the tracking experiment. Reported coefficients come from a regression discontinuity design using Imbens and Kalyanaraman (2012)'s optimal bandwidth. Standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 22: Impact estimations of peers characteristics on PSU performance

Dep. Var.: PSU Score		Math (2)	Science (3)
Panel A: IV Estimation			
Peers Mean Baseline Score	0.00623	0.133*	0.188
	(0.0804)	(0.0774)	(0.140)
Peers Standard Deviation in Baseline	-0.0101	-0.209	-0.239
	(0.176)	(0.168)	(0.397)
Observations	548	561	357
Panel B: First Stage Regressions Peers Mean Baseline Score			
Low Ability Class	-96.413***	-116.900***	-102.452***
	(2.396)	(1.568)	(0.821)
Mixed Ability Class	-46.885***	-59.056***	-53.415***
	(1.955)	(1.245)	(0.693)
F-Test	809.93	2782.36	7801.87
Panel C: First Stage Regressions Peers Standard Deviation in Baseline			
Low Ability Class	5.062***	-34.731***	-28.405***
	(0.614)	(1.429)	(0.858)
Mixed Ability Class	30.759***	8.839***	0.989***
	(0.501)	(1.135)	(0.725)
F-Test	2586.03	713.09	881.25

Note: Reported coefficients come from IV regressions. All regressions control for baseline exam scores with quadratic and cubic polynomials, and dummies for time blocks within which students were assigned to tracked or non-tracked classes. Low- (high-) ability class includes only students below (above) the median PSU baseline score for each subject. Mixed ability classes include every type of students. In Panels B and C, assignment to high-ability class is omitted. Standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix A

Table A1: Heterogeneous impacts by baseline characteristics

Dep. Var.: PSU Score	Language (1)	Math (2)	Science (3)	History (4)
T X Predicted PSU Math Score	-0.0364**	0.103***	-0.0830**	0.0109
	(0.0182)	(0.0162)	(0.0409)	(0.0304)
T X Years of Education	0.379	-0.153	6.680***	-0.955
	(0.892)	(0.801)	(1.835)	(1.203)
T X High School GPA	-11.31	-1.641	-22.94	-21.34**
	(7.723)	(6.942)	(16.66)	(9.811)
T X School's SIMCE	-0.0589	-0.0471	-0.0963	-0.0524
	(0.124)	(0.111)	(0.258)	(0.163)
T X Books in Household	3.886**	0.289	3.765	4.209*
	(1.905)	(1.713)	(4.213)	(2.418)
T X Perception of Control Over	1.938	-2.628	-5.110	10.21
	(5.108)	(4.588)	(11.91)	(6.482)

Note: T is the estimate of the impact of offering a scholarship. Each row reports the coefficient of the interaction between treatment and baseline variables. Reported coefficients come from ordinary least square regressions. All regressions include dummies for the interaction of treatment with cohorts, high-performing schools, and the interaction. Additionally they include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. They also include controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. Results for science and history are estimated correcting for self-selection into the test. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A 2: Impact estimations for the scholarship experiment on PSU number of correct answers

Dep. Var.:	Language		Math		Scienc		History	
Correct R...	(1)	(2)	(3)	(4)	e (5)	(6)	(7)	(8)
Panel A - Impact of Offering a Scholarship for Preu UC								
T	0.402	0.550	1.077**	1.518*	0.765	1.004	0.747	0.993
	(0.390)	(0.672)	(0.465)	(0.804)	(0.713)	(1.237)	(1.133)	(2.001)
T x C2		1.551		0.565		-0.0722		7.828***
		(1.015)		(1.212)		(1.803)		(2.786)
T x HQ		-1.549		-0.369		0.291		-2.640
		(0.971)		(1.161)		(1.772)		(2.906)
T x C2 x HQ		-0.237		-2.345		-1.703		-10.10***
		(1.298)		(1.550)		(2.345)		(3.907)
Observations	1,946	1,946	1,946	1,946	1,235	1,235	773	773
R-squared	0.651	0.653	0.763	0.764	0.693	0.693	0.596	0.602
Impacts by Subsamples								
C1, LQ		0.550		1.518*		1.004		0.993
C1, HQ		-0.999		1.149		1.295		-1.647
C2, LQ		2.101**		2.083**		0.932		8.821
C2, HQ		0.315		-0.631		-0.4802		-3.919
Panel B - Impact of Attending a Top Preuniversitario								
TP	0.792	0.321	2.793**	2.696*	2.316	2.361	3.478	3.77
	(1.083)	(1.359)	(1.272)	(1.572)	(2.006)	(2.405)	(2.721)	(3.878)
TP x C2		6.446**		2.16		1.258		16.62***
		(2.584)		(2.996)		(4.721)		(6.305)
TP x HQ		-4.447*		-0.179		0.32		-11.2
		(2.498)		(2.897)		(4.355)		(7.634)
TP x C2 x HQ		0.318		-4.182		-4.204		-14.09**
		(2.235)		(2.593)		(3.952)		(6.27)
Observations	1827	1827	1826	1826	1158	1158	724	724
Impacts by Subsamples								
C1, LQ		0.321		2.696*		2.361		3.770
C1, HQ		-4.126*		2.517		2.681		-7.43
C2, LQ		6.767***		4.856*		3.619		20.39
C2, HQ		2.638		0.495		-0.265		-4.9

Note: This table shows the scholarship experiment impact on the number of correct responses on the PSU exam. T is the estimate of the impact of offering a scholarship and TP of attending a top *preuniversitario*. C1 and C2 correspond to cohorts of years 2010 and 2011, respectively. HQ and LQ represent high and low school quality level defined by SIMCE test scores. Each row reports the coefficient of the interaction between treatment and baseline variables. Reported coefficients come from OLS regressions for Panel A (ITT estimates), and from IV regressions for Panel B (LATE estimates). All regressions include dummies for cohorts and high-performing schools, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. Results for science and history are estimated correcting for self-selection into the test. The impact by subsample shows the T and TP estimates when the sample is restricted by the interaction of cohorts and high school qualities.
* significant at 10%, ** significant at 5%, *** significant at 1%.

Table A 3: Weights for the average PSU score and wage indexes

Dep. Var.	Average PSU	Log (Wage)
GPA	0.297***	-0.000
	(0.091)	(0.001)
History PSU not taken	57.815**	0.254
	(26.871)	(0.372)
PSU Score, History	0.092**	0.001
	(0.045)	(0.001)
Science PSU not taken	99.545***	0.129
	(30.477)	(0.379)
PSU Score, Science	0.167***	0.001
	(0.054)	(0.001)
PSU Score, Language	0.195***	0.001*
	(0.048)	(0.001)
PSU Score, Math	0.389***	0.004***
	(0.056)	(0.001)
Observations	504	628
R-squared	0.836	0.589

Note: This table generates the weight of the indexes of the average PSU score and wage predictions. GPA indicates students' school grades in PSU score scale. Wage is the logarithm of the average wage, adjusted by the employment rate, of the graduates from the specific career and university attended by the student, four years after graduation. Average PSU is the average of language and math PSU scores obtained by students who attended the same career and institution as the students in the experiment. Reported coefficients come from ordinary least square regressions. All regressions include cohort and high school quality dummies, and controls for gender, school, and household characteristics, academic performance and dedication, and personality traits. Standard errors are in parentheses. * significant at 10%, ** significant at 5%, *** significant at 1%.

Appendix B

1. Survey instruments

Both baseline and follow-up surveys were conducted as measurement instruments (see text of the paper). We attach a sample of the follow-up survey of 2011 and 2012 in different files named "Annex 1-Encuesta de Seguimiento 2011, Preu.docx" and "Annex 1-Encuesta de Seguimiento 2012, Preu.docx," respectively.

2. Power calculations

July 26, 2010

Power has been calculated for three outcomes:

1 PSU test score
2 Entrance into high-quality higher education
3 Returns on education (monthly salary at age 23)

All power calculations assume the following:

- Treatment group size = 627[35]
- Control group size = 489
- Power = 80%
- Alpha = 5%
- One sided t-tests

Correction for imperfect compliance
For the two continuous outcomes (PSU test score and monthly salary), we provide a number of different scenarios, modifying two key parameters:

- Percentage of control group students receiving similar-quality classroom *preuniversitario* programs (10%; 25%; 50%)[36]
- Percentage of treatment group turning down scholarship, but ultimately receiving similar-quality *preuniversitario* course elsewhere (10%; 25%; 50%).

The second parameter is used to calculate the percentage of treatment students actually treated based on the actual takeup of our scholarship (78%) and an estimated dropout rate (once the program has started) of 10 per cent.

Methodology
Minimal detectable effects were computed using the Stata *sampsi* command. The two continuous variables were corrected for imperfect compliance using a multiplier as described in Duflo, Glennerster and Kremer (2007) (*Handbook of Development Economics*). Do-file and output are available on request.

[35] This is the size of the ITT group – those to whom we offered the scholarships
[36] A survey done by a Chilean newspaper estimates 45–65% of students attend some type of *preuniversitario*, but not all classroom *preuniversitario* programs are as intensive or of the same quality as Preu UC. Many *preuniversitario* programs are run by high schools and require fewer classroom hours with poorer-quality professors.

Outcome: PSU Test Score
Baseline value: 478 (from baseline test scores)

Minimal detectable effects for PSU test score (parenthesis indicates standardized effect size) *Attrition Assumption: 5%*				
		Takeup in control group		
		10%	**25%**	**50%**
Percentage of those turning down scholarship that eventually receive similar-quality *preuniversitario* program elsewhere	**10%**	15.8 (.25)	20.9 (.33)	44.5 (.69)
	25%	15.1 (.24)	19.7 (.31)	39.2 (.61)
	50%	14.1 (.22)	17.9 (.28)	32.8 (.51)

Outcome: Percentage entering into high-quality university
Baseline value: 12.3% [37]

Minimal detectable effect for entrance into high-quality university	
Attrition	**Minimal Detectable Effect**
5%	5.7 p.p.
10%	5.9 pp
15%	6.0 pp

Note: These calculations do not take into consideration contamination in control group and imperfect compliance in treatment group.

Outcome: Monthly salary at age 23
Baseline Value: 242,343 Chilean Pesos (Source: CASEN 2008)

Minimal detectable effects for monthly salary (in Chilean pesos) at age 23 (parenthesis indicates standardized effect size) *Attrition Assumption: 20%*				
		Takeup in control group		
		10%	**25%**	**50%**
Percentage of those turning down scholarship that eventually receive same-quality *preuniversitario* program elsewhere	**10%**	58,539 (.27)	77,151 (.36)	164,111 (.76)
	25%	55,871 (.26)	72,582 (.33)	144,731 (.67)
	50%	51,925 (.24)	66,061 (.30)	120,930 (.56)

[37] Entrance into a high-quality university is defined by whether the student has enrolled on a major that captures the best 20 per cent of students by PSU score. Many of these spots are captured by students from expensive private schools, which explains why our baseline value is less than 20 per cent.

References

Alesina, A. and Angeletos, G.M., 2005. Fairness and redistribution. American Economic Review, 95(4), pp.960–80.

Allalouf, A. and Ben-Shakhar, G., 1998. The effect of coaching on the predictive validity of scholastic aptitude tests. Journal of Educational Measurement, 35(1), pp.31–47.

Angrist, J.D. and Lang, K., 2004. Does school integration generate peer effects? evidence from Boston's Metco program. American Economic Review, 94(5), pp.1,613–34.

Barrow, L. and Rouse, C.E., 2005. Do returns to schooling differ by race and ethnicity? American Economic Review, 95(2), pp.83–87.

Becker, B.J., 1990. Coaching for the scholastic aptitude test: further synthesis and appraisal. Review of Educational Research, Fall 60, pp.373–471.

Blossfeld, H.-P. and Shavit, Y., 1993. Persisting barriers: changes in educational opportunities in thirteen countries. In: Y. Shavit and H.-P. Blossfeld, eds, 1993. Persistent inequality: changing educational attainment in thirteen countries. Boulder, CO: Westview Press, pp.1–24.

Blundell, R., Richard, L., Goodman, A. and Reed, H., 2000. The returns to higher education in Britain: evidence from a British cohort. Economic Journal, 110(461), pp.F82–99.

Briggs, D., 2001. The effect of admission test preparation: evidence from NELS-88. Chance, 14(1), pp.10–18.

Buchmann, D., Congdron, D. and Roscigno, V., 2010. Shadow education, American style: test preparation, the SAT and college enrollment. Social Forces, 89(2), pp.435–61.

Calonico, S., Cattaneo, M. and Titiunik, R., 2014. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica, 82(6), pp.2,295–326, November 2014.

Card, D., 2001. Estimating the return to schooling: progress on some persistent econometric problems. Econometrica, 69(5), pp.1,127–60.

CASEN, 2011. Encuesta de Caracterización Socioeconómica Nacional (CASEN) Ministerio de Desarrollo Social, Chile.

Cellini, S. and Chaudhary, L., 2012. The labor market returns to a for-profit college education. NBER Working Papers 18343, National Bureau of Economic Research, Inc.

CNED, 2012. Índices 2012. Consejo Nacional de Educación, Gobierno de Chile.

CONFECH, 2011. Petitiorio CONFECH. Confederación de Estudiantes de Chile. June 26.

DEMRE, 2012. Publicaciones. Departamento de Evaluación, Medición y Registro Educacional, Universidad de Chile.

Duflo, E., Glennerster, R. and Kremer, M., 2007. Using Randomization in Development Economics Research: A Toolkit. In: T.P. Schults and J. Strauss, eds, Handbook of Development Economics, Elsevier Science Ltd.: North Holland, Vol. 4

Duflo, E., Dupas, P. and Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. American Economic Review, 101(5), pp.1,739–74.

Ellwood, D. and Kane, T., 2000. Who is getting a college education? Family background and the growing gaps in enrollment. In: S. Danziger and J. Waldfogel, eds, 2000. Securing the future: investing in children from birth to college. Russell Sage Foundation.

Figueroa, N. and Huepe, M., 2012. La organización industrial de la educación universitaria en Chile. Background study, Grupo Res Pública Chile.

Friedman, W., Kremer, M., Miguel, E. and Thornton, R., 2011. Education as liberation? NBER Working Papers 16939, National Bureau of Economic Research, Inc., April.

Hastings, J.S., Neilson, C. A. and Zimmerman, S. D., 2013. Are some degrees worth more than others? Evidence from college admission cutoffs in Chile. NBER Working Papers 19241, National Bureau of Economic Research, Inc.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica, 47(1), pp.153–61.

Hoxby, C., 2000. Peer effects in the classroom: learning from gender and race variation. NBER Working Paper 7867, National Bureau of Economic Research.

Hoxby, C.M., 2009. The changing selectivity of American colleges. Journal of Economic Perspectives, 23(4), pp.95–118.

Hsieh, C.T. and Urquiola, M., 2006. The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. Journal of Public Economics, 90(8–9), pp.1,477–503.

Imbens, G. and Kalyanaraman, K., 2012. Optimal bandwidth selection in regression discontinuity designs. Review of Economic Studies, 79(3), pp.933–59.

Katz, L.F. and Murphy, K.M., 1992. Changes in relative wages, 1963–1987: supply and demand factors. Quarterly Journal of Economics, 107(1), pp.35–78.

Kling, J.R., 2001. Interpreting Instrumental Variables Estimates of the Returns to Schooling. Journal of Business and Economic Statistics, 19(3), pp. 358–64.

Lang, K. and Weinstein, R., 2012. Evaluating student outcomes at for-profit colleges. NBER Working Papers 18201, National Bureau of Economic Research, Inc.

Lavy, V., Paserman, M.D. and Schlosser, A., 2012. Inside the black box of ability peer effects: evidence from variation in the proportion of low achievers in the classroom. Economic Journal, 122(559), pp.208–37.

MINEDUC, 2012. Informe Nacional de Resultados del SIMCE. Ministerio de Educación, Gobierno de Chile.

MINEDUC, 2013. Educación Media Técnico Profesional: Hallazgos del seguimiento de una generación. Centro de Estudios, MINEDUC, Gobierno de Chile.

Núñez, J. and Millán, I., 2002. ¿Pueden Mejorar su PAA los Alumnos de Escasos Recursos? Evidencia Experimental. Latin American Journal of Economics, 39(116), pp.5–25.

Piketty, T., 1995. Social mobility and redistributive politics. Quarterly Journal of Economics, 110(3), pp.551–84.

Pop-Eleches, C. and Urquiola, M., 2013. Going to a better school: effects and behavioral responses. American Economic Review, 103(4), pp.1,289–324.

Powers, D. E. and Rock, D. A., 1999. Effects of coaching on SAT I: reasoning test scores. Journal of Educational Measurement, 36(2), pp.93–118.

Reyes, L., Rodríguez, J. and Urzúa, S.S., 2013. Heterogeneous economic returns to postsecondary degrees: evidence from Chile. NBER Working Papers 18817, National Bureau of Economic Research, Inc., February.

Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Quarterly Journal of Economics, 116(2), pp.681–704.

Wachtel, P., 1975. The returns to investment in higher education: another view. In: Education, income, and human behavior, NBER Chapters, pp.151–70. National Bureau of Economic Research, Inc.

Winston, G. and Zimmerman, D. 2004. Peer effects in higher education. In: College choices: the economics of where to go, when to go, and how to pay for it, NBER Chapters, pp. 395–424. National Bureau of Economic Research, Inc.

Zwick, R., 2012. The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. Pensamiento Educativo, 49(2), pp.23–30

Publications in the 3ie Impact Evaluation Report Series

The following reports are available from http://www.3ieimpact.org/en/publications/3ie-impact-evaluation-reports/3ie-impact-evaluations/

Sustainability of impact: dimensions of decline and persistence in adopting a biofortified crop in Uganda 3ie Impact Evaluation Report 35. McNiven, S, Gilligan, DO and Hotz, C (2016)

Micro entrepreneurship support programme in Chile, 3ie Impact Evaluation Report 40. Martínez, CA, Puentes, EE and Ruiz-Tagle, JV (2016)

Thirty-five years later: evaluating the impacts of a child health and family planning programme in Bangladesh, 3ie Impact Evaluation Report 39. Barham, T, Kuhn, R, Menken, J and Razzaque, A (2016)

Can egovernance reduce capture of public programmes? Experimental evidence from India's employment guarantee, 3ie Impact Evaluation Report 31. Banerjee, A, Duflo, E, Imbert, C, Mathew, S and Pande, R (2015)

Smallholder access to weather securities in India: demand and impact on production decisions, 3ie Impact Evaluation Report 28. Ceballos, F, Manuel, I, Robles, M and Butler, A (2015)

What happens once the intervention ends? The medium-term impacts of a cash transfer programme in Malawi, 3ie Impact Evaluation Report 27. Baird, S, Chirwa, E, McIntosh, C, and Özler, B (2015)

Validation of hearing screening procedures in Ecuadorian schools, 3ie Impact Evaluation Report 26. Muñoz, K, White, K, Callow-Heusser, C and Ortiz, E (2015)

Assessing the impact of farmer field schools on fertilizer use in China, 3ie Impact Evaluation Report 25. Burger, N, Fu, M, Gu, K, Jia, X, Kumar, KB and Mingliang, G (2015)

The SASA! study: a cluster randomised trial to assess the impact of a violence and HIV prevention programme in Kampala, Uganda, 3ie Impact Evaluation Report 24. Watts, C, Devries, K, Kiss, L, Abramsky, T, Kyegombe, N and Michau, L (2014)

Enhancing food production and food security through improved inputs: an evaluation of Tanzania's National Agricultural Input Voucher Scheme with a focus on gender impacts, 3ie Impact Evaluation Report 23. Gine, X, Patel, S, Cuellar-Martinez, C, McCoy, S and Lauren, R (2015)

A wide angle view of learning: evaluation of the CCE and LEP programmes in Haryana, 3ie Impact Evaluation Report 22. Duflo, E, Berry, J, Mukerji, S and Shotland, M (2015)

Shelter from the storm: upgrading housing infrastructure in Latin American slums, *3ie Impact Evaluation Report 21*. Galiani, S, Gertler, P, Cooper, R, Martinez, S, Ross, A and Undurraga, R (2015)

Environmental and socioeconomic impacts of Mexico's payments for ecosystem services programme, *3ie Impact Evaluation Report 20.* Alix-Garcia, J, Aronson, G, Radeloff, V, Ramirez-Reyes, C, Shapiro, E, Sims, K and Yañez-Pagans, P (2015)

A randomised evaluation of the effects of an agricultural insurance programme on rural households' behaviour: evidence from China, 3ie Impact Evaluation Report 19. Cai, J, de Janvry, A and Sadoulet, E (2014)

Impact of malaria control and enhanced literacy instruction on educational outcomes among school children in Kenya: a multi-sectoral, prospective, randomised evaluation, 3ie Impact Evaluation Report 18. Brooker, S and Halliday, K (2015)

Assessing long-term impacts of conditional cash transfers on children and young adults in rural Nicaragua, *3ie Impact Evaluation Report 17*. Barham, T, Macours, K, Maluccio, JA, Regalia, F, Aguilera, V and Moncada, ME (2014)

The impact of mother literacy and participation programmes on child learning: evidence from a randomised evaluation in India, 3ie Impact Evaluation Report 16. Banerji, R, Berry, J and Shortland, M (2014)

A youth wage subsidy experiment for South Africa, 3ie Impact Evaluation Report 15. Levinsohn, J, Rankin, N, Roberts, G and Schöer, V (2014)

Providing collateral and improving product market access for smallholder farmers: a randomised evaluation of inventory credit in Sierra Leone, 3ie Impact Evaluation Report 14. Casaburi, L, Glennerster, R, Suri, T and Kamara, S (2014)

Scaling up male circumcision service provision: results from a randomised evaluation in Malawi, 3ie Impact Evaluation Report 13. Thornton, R, Chinkhumba, J, Godlonton, S and Pierotti, R (2014)

Targeting the poor: evidence from a field experiment in Indonesia, 3ie Impact Evaluation Report 12. Atlas, V, Banerjee, A, Hanna, R, Olken, B, Wai-poi, M and Purnamasari, R (2014)

An impact evaluation of information disclosure on elected representatives' performance: evidence from rural and urban India, 3ie Impact Evaluation Report 11. Banerjee, A, Duflo, E, Imbert, C, Pande, R, Walton, M and Mahapatra, B (2014)

Truth-telling by third-party audits and the response of polluting firms: Experimental evidence from India, 3ie Impact Evaluation Report 10. Duflo, E, Greenstone, M, Pande, R and Ryan, N (2013)

No margin, no mission? Evaluating the role of incentives in the distribution of public goods in Zambia, *3ie Impact Evaluation Report 9*. Ashraf, N, Bandiera, O and Jack, K (2013)

Paying for performance in China's battle against anaemia, *3ie Impact Evaluation Report 8*. Zhang, L, Rozelle, S and Shi, Y (2013)

Social and economic impacts of Tuungane: final report on the effects of a community-driven reconstruction programme in the Democratic Republic of Congo, *3ie Impact Evaluation Report 7*. Humphreys, M, Sanchez de la Sierra, R and van der Windt, P (2013)

The impact of daycare on maternal labour supply and child development in Mexico, *3ie Impact Evaluation Report 6*. Angeles, G, Gadsden, P, Galiani, S, Gertler, P, Herrera, A, Kariger, P and Seira, E (2014)

Impact evaluation of the non-contributory social pension programme 70 y más in Mexico, *3ie Impact Evaluation Report 5.* Rodríguez, A, Espinoza, B, Tamayo, K, Pereda, P, Góngora, V, Tagliaferro, G and Solís, M (2014)

Does marginal cost pricing of electricity affect groundwater pumping behaviour of farmers? Evidence from India, 3ie Impact Evaluation Report 4. Meenakshi, JV, Banerji, A, Mukherji, A and Gupta, A (2013)

The GoBifo project evaluation report: Assessing the impacts of community-driven development in Sierra Leone, *3ie Impact Evaluation Report 3*. Casey, K, Glennerster, R and Miguel, E (2013)

A rapid assessment randomised-controlled trial of improved cookstoves in rural Ghana, *3ie Impact Evaluation Report 2*. Burwen, J and Levine, DI (2012)

The promise of preschool in Africa: A randomised impact evaluation of early childhood development in rural Mozambique, *3ie Impact Evaluation Report 1*. Martinez, S, Naudeau, S and Pereira, V (2012)

This study in Chile evaluated the impact of providing scholarships to secondary school students from low- and middle-income households to help them attend high-quality training institutions to prepare for standardised entrance tests for college admission.

Researchers assessed the impact over two years. In the first year, they found non-significant impact on the test scores and higher education entry. However, in the second year, there was significant impact recorded among students from low-performing schools. There were several school days lost due to student protests during this year. The study showed that the scholarships helped students use the training to substitute for preparation in school. It also found that the opportunities that high school students received affected their attitudes and opinions on increasing or decreasing public expenditure to subsidise higher education in Chile.

Impact Evaluation Series

International Initiative for Impact Evaluation
202-203, Rectangle One
D-4, Saket District Centre
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org