# Screening and Recruiting Talent At Teacher Colleges using Pre-College Academic Achievement

Sebastian Gallegos
IDB

Christopher Neilson
Princeton University and NBER

Franco Calle
Princeton University

September 4, 2019

In this paper, we study policies that use pre-college academic achievement to restrict or incentivize entry to teacher-colleges. We use historical records of college entrance exam scores since 1967 in Chile and administrative data on the population of teachers to document a robust positive and concave relationship between several measures of teacher productivity and teachers' own pre-college academic achievement. We then study the effectiveness of two policies that used pre-college achievement to recruit or screen out students entering teacher-colleges. Using a regression discontinuity design based on the government's recruitment efforts, we evaluate the effectiveness of targeted scholarships at shifting career choices of high achieving students as well as the effect on the overall stock of teachers predicted effectiveness. We then study the effects of a recent screening policy that forces teacher colleges to exclude below-average students. We retroactively simulate this policy rule and evaluate its success at screening out low performing teachers. We compare this benchmark policy rule to a series of potential data-driven policy rules and we find that even simple screening policies can identify a significant portion of ex-post low performing teachers. Compared to the benchmark policy rule, a more flexible data-driven model is able to reduce the amount of ex-post high performing teachers excluded while maintaining the number of low performing teachers screened out. In both policies studied, screening low performing students is more effective than targeting recruitment efforts to only very high achieving students. These findings suggest that the combination of better data and flexible prediction methods can be used to implement practical screening policies in some contexts.

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE.

*Gallegos: Inter-American Development Bank, 1300 New York Avenue, Washington, DC. Neilson (corresponding author) and Calle: Department of Economics, Princeton University, 20 Washington Road, Princeton, NJ 08544. Email: cneilson@princeton.edu. We thank the DEMRE and Ministry of Education (MINEDUC) of the government of Chile for access to the sources of data and help locating and accessing data records. We thank Macarena Alvarado and Alvaro Carril for contributions to early versions of this paper and seminar participants at The University of Chicago, Princeton's Education Research Section, the Workshop on Education and Child Development, University of Stavanger, Norway, the Education Seminar at the Inter-American Development Bank, and the APPAM and SREE annual meetings. Gallegos also thanks the Center for Studies of Conflict and Social Cohesion (CONICYT/FONDAP/15130009). All errors are our own, and the views expressed in this paper are those of the authors and should not be attributed to the Inter-American Development Bank or any other institution. Online Appendix available at https://christopherneilson.github.io/work/teacher-quality.html.

# 1 Introduction

A common policy objective of governments all over the world is to increase the productivity of the stock of teachers. One set of policies looks to increase the effectiveness of current teachers through incentives, increasing their effort through accountability measures and rewards. For instance, Hoxby and Leigh (2004) documented that talented teachers where leaving the profession due to the compression of wages and, **?** who finds that a policy based on rewards, improved effective teaching. These policies typically require measuring teacher effectiveness carefully and tracking it over time to write optimal incentive compatible contracts making them difficult to rely solely on incentives to promote effective teachers in many contexts, especially in poorer developing countries. A second complementary set of policies looks to promote the recruitment of talented individuals into the teaching profession. Policies such as college tuition scholarships make attending teacher college more attractive and screening entry policies can exclude candidates from teaching colleges that do not meet minimal standards. Designing both of these policies efficiently depends crucially on pre-college academic achievement being an important determinant of how effective (or not) a student will be as a teacher in the future. In the US context there is consensus that predicting talented teachers' ex-ante is notoriously difficult, making it also impractical to design systems relying solely on high powered benefits to recruit talented individuals or minimal standards to exclude low performing candidates. In developing countries the relevant margin of quality seems to be closer to whether teachers show up for work (Kremer et al. (2005), Duflo et al (2012), Muralidharan et al. (2014), though some studies suggest that pay for performance schemes could work for those already in the teacher workforce (**??**).

In this paper we study policies that use pre-college academic achievement to 1) screen out applicants that are unlikely to become effective teachers later in their careers and 2) target resources to recruit high achieving students to become teachers. We first show evidence that in the context of a developing country, pre-college academic achievement in school and in college entrance exams is systematically correlated with teacher performance. We use recently digitalized historical data on college entrance exams from 1967 onward, merged with administrative records on the population of teachers in Chile. We document a robust positive and concave relationship between teachers' pre-college academic achievement and a variety of short, medium and long run teacher outcome measures. Short and medium run outcomes include the probability of graduation from teacher college, college exit exams, and employment and wages in schools. Long run measures of productivity include government teacher evaluations, student test scores, school value added as well as own teacher value added. Broadly we find that below average students are systematically associated with lower performance as teachers when their performance is measured up to thirty and forty years later. Our findings suggest that at least in the context of a developing economy such as Chile, own academic readiness can be seen as a necessary condition. College training nor experience seem to be enough undo the initial deficiencies and the relationship between entrance exam and measures of performance.

We then evaluate two policies that look to affect the distribution pre-college academic Achievement of the students entering teaching colleges in the country. The first is a policy implemented to attract high scoring students by offering tuition subsidies at participating teacher colleges. The policy also asked teacher colleges to voluntarily exclude students with below average entrance exam scores if they wanted to be eligible for the program. The second policy which was implemented later, forced all teacher colleges to exclude students with below average college entrance exams with an allowance for students with high ranking GPA. We show that these policies had a significant impact on both individual major choices of students as well as an effect on the aggregate distribution of predicted teacher quality.

To evaluate the first recruiting policy enacted in 2011, we use a regression discontinuity design to show how high-powered incentives such as scholarships can attract high achieving students to the teaching profession. Choice probabilities for students with scores one standard deviation above the mean jumped 40% at the cutoff but on a small base leading to an increase of 1000 (out of 20000) additional students with high academic achievement. Approximately fifty percent of teacher colleges participated and had to exclude the majority of low performing applicants. This did significantly increase average pre-college academic Achievement by cutting off a significate part of the lower tale of the distribution and as a result, raised predicted teacher effectiveness. Higher achieving students graduated and performed as predicted on exit exams six years later lending credence to the idea the relationships documented above are invariant to this type of policy. However, several years later, tuition was made free at many other options due to a different policy and the same regression discontinuity showed that the effectiveness of the policy was greatly reduced. This result serves as a reminder that recruitment incentives are only as good as the next best option and that high achieving students have many good alternatives so it is harder to move them towards teaching.

We then turn to study a second policy using pre-college academic achievement is a direct screening policy enacted in 2017 that excluded low performing students from entering teaching colleges anywhere independently of whether the college wanted to participate or not. We evaluate this very recent policy by replicating it going backward and identifying who would have been affected and whether they did indeed become ineffective teachers. Partial equilibrium analysis shows that if implemented, these rules would have bound for 25% of students entering teaching colleges in 2016 and would have affected 20% of current teachers, including 87% of the worst performers based on government teacher evaluations. We then evaluate the scope to improve the screening by using standard machine learning methods to determine the predictive content of the information available on academic achievement prior to entry to college. We train a standard model that classifies potential teachers based on the rich set of precollege test information. The results suggest policy makers can use initial characteristics to determine college specific cutoff rules that exclude students with 90% chance of being a low performing teacher. Partial equilibrium analysis shows that if implemented, these rules would have bound for 25% of students entering teaching colleges in 2016 and would have affected 20% of current teachers, including 87% of the worst performers based on government teacher evaluations.

Taken together these results indicate that policies that use pre-college academic achievement can be useful to increase the predicted quality of teachers. It is also the case that data driven adjustments to the policies could make them more effective. In addition, given we find that predicting lower performing teachers is easier than high performers, screening policies seem the most viable policy in this context, possibly complemented with targeted subsidies to students in a broad "competent category" instead of chasing after the really high achieving students.

These results are important because they have direct policy implications. If teacher effectiveness or the lack of it is possible to predict early on, policies could focus resources on recruiting and retaining the most promising candidates and filtering out applicants who are more likely to be ineffective teachers later on. This is particularly important because teacher labor markets are known to be inefficient, misallocation of talent can be widespread and in many cases, there is limited scope to sideline or retrain ineffective teachers once they are in the system, especially in the public sector. The results presented in this paper highlight avenues for further research in an increasingly data rich environment where prediction is a key input to policy design. While entrance exam scores, transcripts and GPA remain an imperfect measure of teacher quality, they are broadly available and can be used to target resources and design policy at zero marginal cost. In many developing countries this is the only option they have to design policy as implementing

the infrastructure needed to implement repeated student tests or teacher classroom evaluations regularly is not available. Our findings suggest that at least in the context of low to middle income countries such as Chile, resources that look to subsidize teacher training should be targeted towards prospective teachers that have a minimal level of baseline academic achievement. Own academic readiness can be seen as a necessary condition (maybe not sufficient). These results are related to recent work in developing countries has shown that teacher value added is important and varies a lot. For example Bau and Das (2018) show evidence on teacher value added from Pakistan finding important misallocation of pay in the public sector and in contrast to our results that teacher quality value added is not explained by teacher observable characteristics. Araujo et al. (2016) estimate teacher value added randomly assigning students to teachers and find significant variation in teacher quality and that observational measures of quality were correlated with estimated value added. Similarly, to Bau and Das (2018), they do not find teacher characteristics to be correlated with teacher value added. The results presented in this paper seem potentially at odds with this evidence and that from the US (Rockoff, 2004; **?**). However, there are important differences worth mentioning is the empirical exercise. In our case we are using rich pre-college academic achievement for the population of teachers that is not available to researchers in other contexts. Another important thing to note is that we find most of the correlation comes from very low achieving students who become teachers and this margin may not be relevant for more developed countries.

## 2 Context, Data and Descriptive Statistics

### 2.1 Teachers In Chile

There are about 164,000 classroom teachers in Chile (MINEDUC, 2019) who teach almost 3 million students in grades 1st to 12th.[1] The country has reached almost universal levels of educational enrollment in primary (99%) and secondary school (92%), which are well above other Latin American countries (OECD, 2009) and similar to developed countries. Chile is in the advanced stages of a demographic transition, with low fertility and mortality rates, and relatively high life expectancy (World Bank, 2011). Consequently, enrollment in primary and secondary education has plateaued and even showed a slight decrease over the last ten years (from 3.1 million in 2008 to 2.9 million in 2018). In the meantime, the number of classroom teachers has increased importantly (from 125,000 in 2008 to 164,000 in 2018), which has led to a reduced student-teacher ratio (from 25 to 18) that is close to the levels displayed by OECD countries (World Bank, 2013). In addition, teacher absenteeism is low (5%, see Paredes et al. (2015)) unlike other countries in earlier stages of development (19 % on average in Bangladesh, Ecuador, India, Indonesia, Peru and Uganda (Chaudhury et al., 2006). With enough teachers in the classrooms, the policy focus switched in the last ten years to bring more qualified individuals to the teaching profession.

Attracting more skilled individuals to be teachers is a challenge because, among other factors, teachers are paid less (Mizala and Nopo, 2016) compared to other professionals.[2]

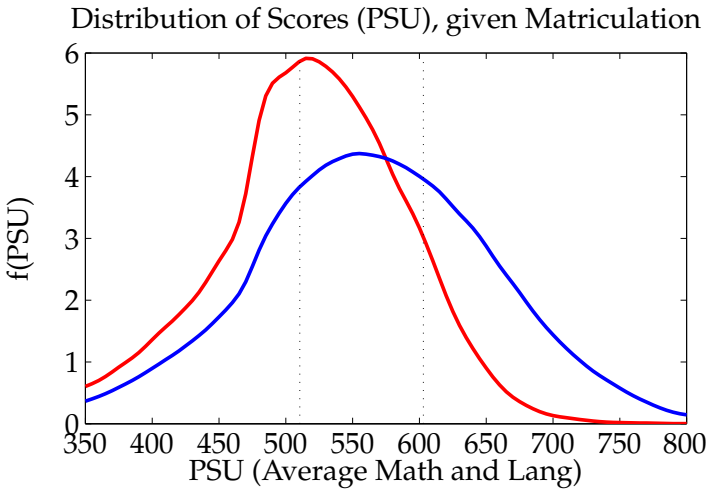
Consistently, we know from the related literature that college graduates with higher college entrance scores are less likely to enter teaching (Manski, 1985; Hanushek and Pace, 1995; Vegas

[1] Teachers work in the different types of schools which differ in their funding and administration. Public schools are funded and administered by the government; voucher schools are funded mainly with public funds but administered by privates; and private schools are both funded and administered privately.

[2] Mizala and Nopo (2016) show that teachers are underpaid in Latin America compared to other professionals and technicians, after controlling by observable characteristics linked to productivity. In Chile in particular, the underpayment for teachers was about 18% in 2007.

et al., 2001), and Chile is no exception. Figure 1 shows that students enrolled in fields other than education (engineering, law, medicine, etc.) scored about 1 standard deviation (SD) above the national mean (600 points) in the college entrance exam, while teacher college students scored 0.1 SD above the national mean. In addition, we computed that the scores from teacher colleges have been declining over time; in 1995 students from teacher colleges scored 0.3 SD over the national mean. This pattern is similar to the evidence for the US (Bacolod, 2006; Corcoran et al., 2004; Podgursky et al., 2004; Hoxby and Leigh, 2004).

**Figure 1:** Freshmen College Exam Scores: Teachers Colleges (red) vs Other Fields (blue), 2007-2010



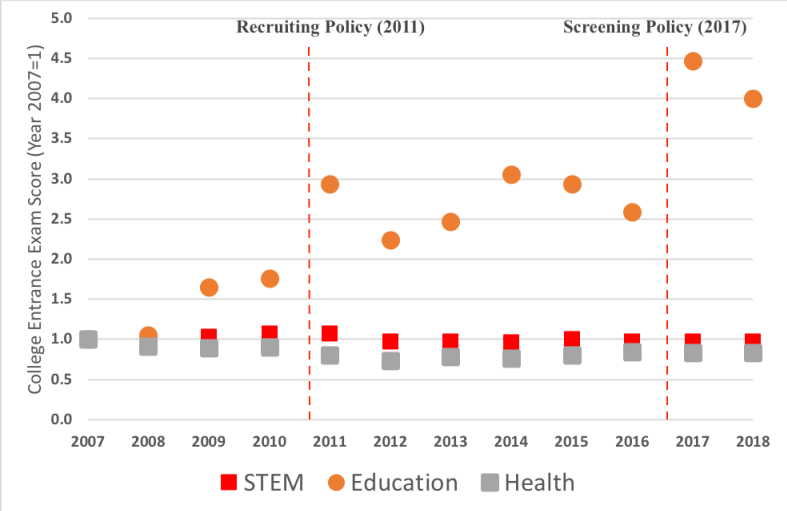

Note: The figure plots the distribution of college entrance exam score for freshmen in education (in red) and any other field (in blue) from 2007 to 2010.

## 2.2 Policies to Recruit Teachers In Chile

The Chilean government implemented in 2011 a scholarship aiming to recruit higher-scoring students to the teaching profession and also to restrict access for the lower-scoring students. The policy, called *Beca Vocacion Profesor (BVP)*, consisted in full tuition subsidies for prospective students who scored about 1SD above the mean in the college entrance exam. The policy also required participating colleges to reject applicants with scores below the mean. In the light of promising results, the government implemented in 2017 a screening policy, imposing new requirements for admissions at all teacher colleges. The government policy implied that applicants to teacher colleges have to have college entrance exam scores at least as high as the 50th percentile of the distribution (or have a high-school GPA in the top 30% of their high school graduating cohort).

Figure 2 shows the evolution of the average college entrance exam for freshmen in education, and is suggestive of the policy effects on scores in 2011 and 2017. The Figure shows data for freshmen in the education, health and STEM fields, from 2007 to 2018, using 2007 as the base year. There is a sharp increase in the scores for freshmen in education from year 2010 to 2011, and another from 2016 to 2017, which coincide with the implementation of the BVP and the new government rule. At the same time, Figure 2 show no variations in the scores achieved by freshmen in STEM or health fields.

**Figure 2:** Relative Freshmen College Exam Scores, by Field




Note: The figure plots the the average college entrance exam score for freshmen in education, health and STEM fields, from 2007 to 2018, using 2007 as the base year.

## 2.3 Data on Pre-College Academic Achievement

The main measure of pre-college academic Achievement that is used in this paper is the college entrance exam scores, similar to the SAT in the U.S. with additional subject tests beyond math and language. The Chilean national college entrance exam is currently called the *Prueba de Selection Universitaria* (PSU). It has been administered once a year since 2004 and prior to that a similar test called *Prueba de Aptitude Academica* had been implemented since 1967 being one of the longest running centralized college assignment systems in the world [3]. Test-takers complete exams in mathematics and language as well as other specialized subjects. The scores are scaled to a distribution with a mean and median of 500 and standard deviation of 110. The exam scores are required to apply to all public universities and most private universities and institutes. The majority of government subsidies have some requirement associated with this test and this will be used in the empirical strategy later in the paper. In this paper, we use data on entrance exams taken since 1967. These data have been collected as a part of the work done in Hastings et al. (2014) where the authors collected digital copies of old books and newspapers, digitalizing test score data back to the first test in 1967.

A second source of pre-college academic Achievement is students GPA and detailed course level transcripts. These are available since 2003 for the population of students. This allows for the creation of both absolute and relative measures of student academic Achievement and comparative advantage. A final source of pre-college academic Achievement are standardized tests taken in school, both in 8th grade as in 10th grade. These are available for the years 2000 onward.

[3] A detailed explanation on the application and enrollment process for the period 1980-2009 is presented in Hastings et al. (2014) and a review comparing centralized systems in the world in Neilson (2019)

## 2.4 Data on Teacher Productivity

During the last decade, Chile has developed important efforts to measure teacher performance. According to a recent review by World Bank Economist, Chile has the most advanced system of teacher performance evaluation in Latin America (Bruns and Luque, 2015). The two most important assessments are exit exams for graduates from teacher colleges and government evaluations for public school teachers. In addition, the Ministry of Education of the Government of Chile (MINEDUC) keeps administrative records with information on teh population of teachers at all schools including private and government subsidized schools since 2003. In spite of the general richness in the data, what is not readilly available are measures of teacher value added because students take standardized tests on irregular grades. To fill this gap, we use value added measures collected for a sample of teachers in (De Gregorio et al., n.d.), where students took tests at the begining and at the end of the academic year to capture value added of the teacher by controling for initial conditions. In (De Gregorio et al., n.d.), students are also surveyed on their perceptions regarding whether their teacher is effective. We use all these sources of information to measure teacher quality and describe them briefly below and Table 17 lists our teacher productivity measures, with their number of observations, age at which the outcome was measured and year of data collection.

Our measures of teacher productivity span over earlier outcomes (at age 23) to longer run outcomes measured more than thirty years later. In particular, our teacher outcomes include short run outcomes such as graduation from teacher colleges and college exit exams; and long run outcomes such as earnings, employment, external classroom teaching evaluations, and student's value added and perception. We describe briefly each dataset below. In our online appendix we describe each measure in detail.

### *2.4.1 Administrative Data Sources*

*Graduation from Teacher Colleges*. We use information from enrollment in teacher colleges for years 2004 to 2009 for about 85K individuals, which we link to graduation records from years 2009 to 2017. We study on time graduation rates (within 5 years after initial enrollment, at 23 years old) and late graduation (up to 8 years after enrollment, at 26 years old).

*Exit Exams*. Our data consists in microdata from all the *INICIA* and *Diagnostica* exam test-takers between 2009 and 2017. The sample consist of about 35K just graduated teachers with scores in a disciplinary knowledge test (e.g., math knowledge for math teachers) and a pedagogical knowledge test (e.g., capacity of explaining concepts in a coherent way). At the time of the exam test-takers were approximately 25 years old on average.

*Government Evaluations*. We gathered information for 63K classroom teachers in public schools, evaluated between 2004 (when the ED started) to 2017. On average, teachers were 40 years old at the time of the evaluation. They have on average 12.5 years of tenure (years working in schools).

*Employment in Schools*. We gathered information for about 240K graduates from teacher colleges in years 1995 to 2017 and merged with the population of employed teachers between 2003 to 2018. We compute whether graduates work ever as teachers and we also study whether they work as teachers 2, 5, 10 years, and 10, 15 and 20 years later, respectively and correlate that with entrance exam scores. The age at employment after ten years and twenty of graduation average 37 and 46 years old respectively.

*Wages in Schools*. Teachers with information on wages are 37 years old, 70 percent female, with 9 years of tenure. Teachers working in public schools are 38 percent of the sample. They benefit from a special labor code (known as the Teacher Statute), and so wages are not expected

to change with productivity (wages grow with tenure in the public sector. However,the voucher sector operates under the regular, more flexible, labor code, and thus teacher wages can be given a market clearing interpretation, associated to productivity. They represent sixty percent of our sample.

The data on employment and wages comes from administrative records on the population of teachers, which is formally reported by the school principal to the MINEDUC at the beginning of each academic year. The data, available for the period 2004-2017, contains 3.4 million records (an average of about 215K teachers per year), which correspond to 370,000 unique teachers. This dataset has information on teachers' demographics (gender, age), specialization level (preschool, primary, secondary) and subject taught (math, language, general, other), as well as tenure (years working as a teacher) and job characteristics (hours of contract, type of contract and starting in 2011, wages).

### 2.4.2 *Collected Data Sources*

*Value Added*. Despite its rich administrative records, Chile does not implement tests designed to measure value added. To overcome this lack of data we implemented those tests for a sample of 9th, 10th and 11th graders and their teachers. The tests measured math skills at the beginning and the end of the academic year, and were specially designed to measure the gain in student achievement during that period.[4]

*Student Perceptions*. As part of the same project described above, we collected information from students' perceptions regarding effective teaching. The perceptions survey contains 38 questions and follows the recommendations of the Measures of Effective Teaching study carried out in the U.S. (Kane and Cantrell, 2010). Questions are categorized eight dimensions of teaching practices and classroom environment: positive culture and learning environment, student understanding checked for and ensured, engaging learning environment, expectations held by teacher, student input and ideas valued, learning fully internalized by students, encouraging and supporting relationships fostered, and classroom participation. We use students' perceptions in a as an additional measure of teacher quality.

## 3 Pre-college Academic Achievement and Measures of Teacher Productivity

In this section we document the systematic correlation between pre-college academic ability and a host of teacher productivity measures, at different moments of their careers. We computed parametric regressions of each outcome on entrance exam scores, allowing for concavity of the relationship. We document the descriptive correlation between pre-college academic ability and teacher productivity measures estimating regressions of each outcome on entrance exam scores. We also describe the empirical relationship showing non-parametric plots leveraging on our large sample sizes.

The general takeaway is that the empirical relationship between pre-college skills as a student and teacher productivity later on is positive and concave. In Table 1 we show the coefficients of 14 separate regressions for different measures of teacher performance on the college entrance

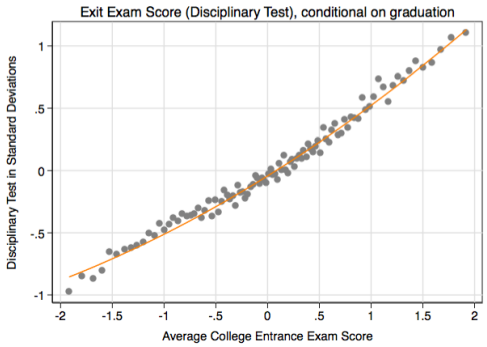
[4]The tests were developed by the Measurement Center of the Catholic University of Chile (MIDE UC) and are called SEPA tests. The name SEPA comes from the acronym in Spanish of Learning Progress Assessment System (*Sistema de Evaluacion de Progreso del Aprendizaje*. In De Gregorio et al. (n.d.) we use these learning gains to study the effectiveness of traditional teachers vis-a-vis "Ensena Chile" teachers. "Ensena Chile" teachers are recent college graduates recruited to teach for two years in low-income communities, resembling the U.S. experience of "Teach for America" teachers.

exam score (in standard deviation units) and its square. The coefficients on scores are positive and significant, and most coefficients on the square are negative. In all cases coefficients should be read as the change in the outcome given one standard deviation (SD) change in the scores.

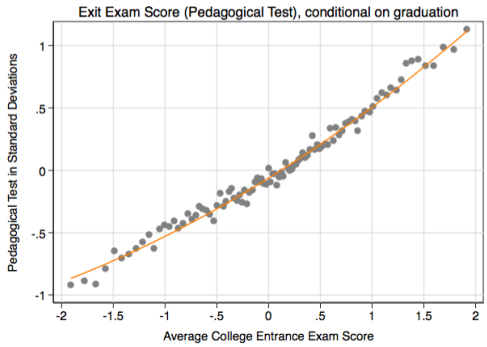
In Figure 3 and Figure 4 we examine early outcomes of students from teacher colleges, like exit exams and graduation rates. We find that college entrance exams scores are positively correlated with the exit examinations college students of pedagogy take before graduating. A one standard deviation on the test scores that current teachers took years ago, is associated to an increase of 0.50 standard deviations on both the disciplinary and pedagogical skills in the exit exams respectively; and an increase in 0.46 and 1.27 standard deviations in writing skills and ICT skills (available for smaller samples). The same pattern can be visualized in Figure 3 for the pedagogical and disciplinary tests. Non parametric plots for ICT and writing tests are shown in the appendix.

**Figure 3:** College Entrance Exam Average and Teacher College Exit Exams

**(a)** Disciplinary Exit Exam



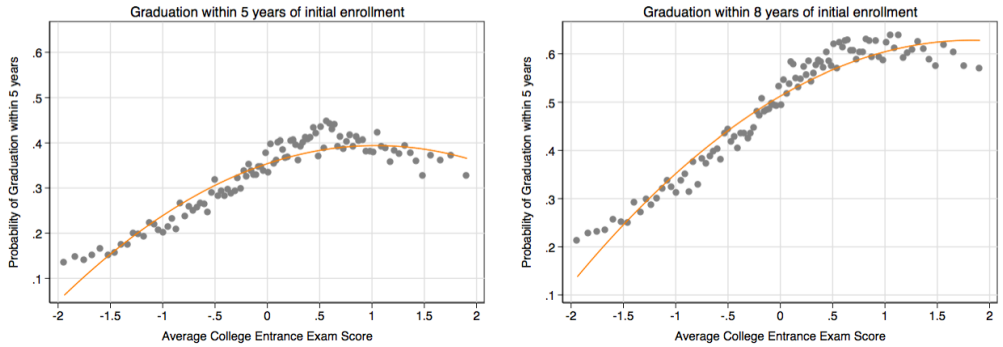

**(b)** General Pedagogy Exit Exam




Note: The figures plot the mean of each exit exam for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data. The data consists in graduates who took the respective exit exam test between years 2009 and 2017. The Figures consider all test-takers with valid scores in the PSU and the Disciplinary test (Figure 3a, $N = 35,355$), and the Pedagogical test (Figure 3b, $N = 33,409$).

**Figure 4:** College Entrance Exam Average and Graduation from Teacher Colleges

**(a)** Graduation within 5 years **(b)** Graduation within 8 years




Note: The figures plot the probability of graduation for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data. The data consists in enrolled students in years 2004 to 2009 who graduated between 2009 and 2017. In both Figures the sample size is of $N = 84,847$.

In terms of graduation, the relationship appears concave for both on-time and late graduation. The results show than an increase in one standard deviation on the college entrance exam scores leads to an increase in the on-time (late) graduation rate from teachers colleges of 7.3 (11.8) percentage points (relative to a baseline graduation rate of 34.7% (47.3%). One hypothesis that might explain the concave relation is that exceptional students might either switch to another career and quit the teaching profession to a more attractive career for them, or they might just have an attractive outside option in the labor market.

Moreover, we show that the pre-college human capital measure is positively related with different measures of productivity once teachers-to-be graduated and entered the labor market. Figure 30 show the bivariate relation between college exam scores taken at 18 and teacher evaluations taken up to 30 years later. The relationship is concave again, suggesting that early scores are not great at predicting good performance as a teacher later on, but they appear to be very good predictors of low performance as a teacher thirty years later. Coefficients in Table 1 show that indeed, an increase of one standard deviation in the scores implies an increase of 0.62 and 0.48 standard deviations on the teacher evaluation score overall and portfolio score respectively.

Figure 6 shows how hourly wages vary with scores, by teachers working in public and private schools. The slope is much steeper for teachers working in the private sector, and rather flat for teachers working in the public sector. Consistently, the coefficients in Table 1 show that a standard deviation increase in scores is associated to 0.26 and 0.45 standard deviations of hourly wages for teachers working in the private and public sector respectively. The magnitude of the coefficient over wages is more prominent for the sample of teachers in the private sector since private schools can move salaries unrestrictedly as teacher productivity changes, the same dynamic does not occur in the public sector where wages are less flexible and determined primarily by the years of service in the public sector which is not a concise measure of productivity.

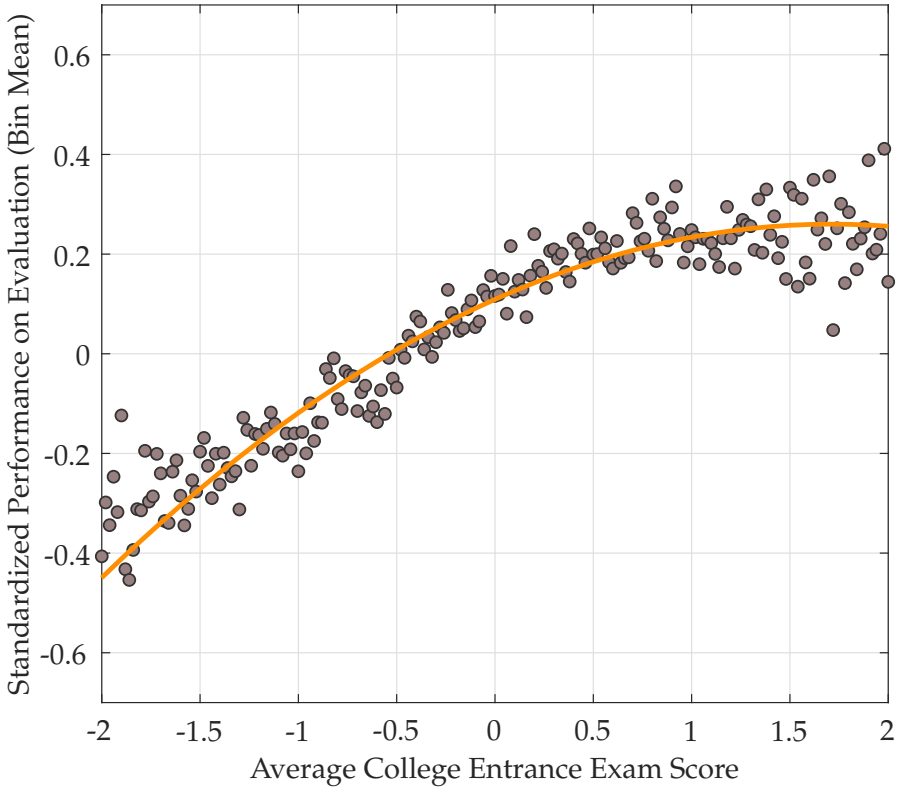
The summary table Table 1 shows that the relationship between pre-college academic achievement and productivity measures appears to be systematically concave. There is a significant share of exceptional students that end up performing worse as teachers than what we would have expected in a linear setting. These results hold when we run non parametric plots for both measures

of productivity in the public and private sector versus PSU scores, as shown in **??**. Some of them might wind up quitting the teaching profession.

The bottom panel of Table 1 shows the results for entrance exams and teacher value added. An increase of one standard deviation in PSU score is associated with an increase of 0.33 standard deviations in math gains over the academic year. The results decomposed by topic (algebra, numbers geometry and statistics) range from 0.29 to 0.43 increases. Results are also suggestive of a concave relationship though the negative coefficients are not significant at conventional levels.

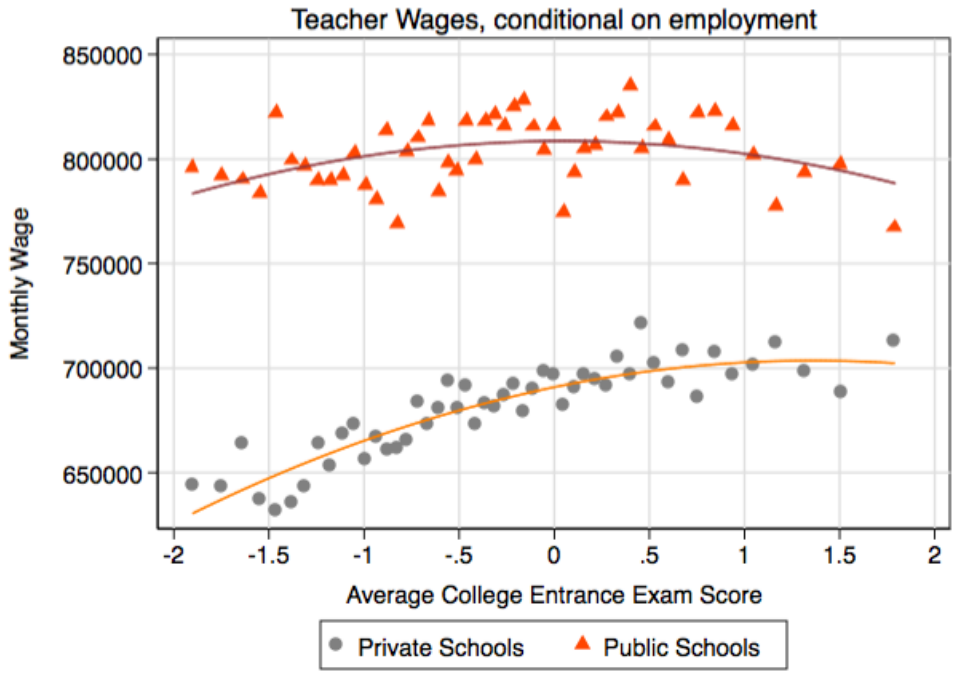
The results for student perceptions also follow the same pattern. We use a factor analysis to produce an index for student perception using the eight categories of teacher effectiveness reported by students. We then regress the index on the teacher's PSU scores and its squared, and do the same for each of the eight categories. The results, presented in **??** show that an of one standard deviation in PSU score is associated to 0.08 SD in students perceptions index and again finding suggestive evidence that the relationship is concave (negative and significant coefficient of -0.03 on the square of the PSU score).

**Figure 5:** Entrance Exams and In-Classroom Observational Evaluation




Note: This figure shows the binned plot of the mean of teacher evaluations conditional on pre-college academic Achievement measured by the average of standardized math and language scores. See Table 1 for a summary of the correlation of pre-college academic Achievement and other measures of teacher productivity.

**Figure 6:** College Entrance Exam Average and Earnings of Teachers




Note: The figures plot the probability of graduation for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data. The data consists in enrolled students in years 2004 to 2009 who graduated between 2009 and 2017. In both Figures the sample size is of $N = 84,847$.

**??** is consistent with concave productivity and shows again a non linear pattern between scores in PSU and the probability of working as a teacher years after graduation. First, an increase of one SD in psu scores increases the likelihood of working as a teacher in 38% percentage points (5 years after graduation) relative to a baseline of 44%. Nevertheless, a significant fraction of teachers in the right tail of the distribution of college preparedness quit the profession by that time.

In the next sections of the paper we study two policies that look to screen out low performing students from teachers colleges or to attract high achieving students to teachers colleges. The first policy called *Beca Vocacion Profesor* is a policy implemented in 2010 which was aimed primarily at recruiting high scoring students into teacher colleges by offering full scholarships, stipends and even paid semesters abroad. Short run effects show significant effects on choices made by students as many enrolled in teacher colleges instead of alternative options. Seven years later, the medium run results on early indicators of teacher productivity strongly suggest that the relationship between pre-college academic achievement and teacher productivity is invariant to the policy studied. The second policy was an across the board mandatory screening policy implemented in 2017 and requires accredited teachers' colleges to only allow students with above average entrance exam test scores into teacher colleges. We study the predicted change in teacher productivity given the policy and simulate the partial equilibrium effects of the policy had it been implemented in the past.

**Table 1:** Summary table: Teaching Performance vs PSU Scores

| Graduation | Years after enrollment | | | |
|---|---|---|---|---|
| | 5 Years | 8 Years | | |
| PSU Score | 0.073*** | 0.118*** | | |
| | ( 0.002 ) | ( 0.002 ) | | |
| (PSU Score)$^2$ | -0.027*** | -0.026*** | | |
| | ( 0.001 ) | ( 0.001 ) | | |
| Observations | [ 84,847 ] | [ 84,847 ] | | |
| Dep. Var. Mean | 0.322 | 0.473 | | |
| Exit Exams | Disciplinary Test | Pedagogy Test | Writing Test | ICT Test |
| PSU Score | 0.509*** | 0.506*** | 0.463*** | 1.27 *** |
| | ( 0.005 ) | ( 0.007 ) | ( 0.007 ) | ( 0.014 ) |
| (PSU Score)$^2$ | 0.043*** | 0.033*** | -0.021*** | -0.07 *** |
| | ( 0.003) | ( 0.311 ) | ( 0.200 ) | ( 0.443 ) |
| Observations | [ 35,355 ] | [ 33,409 ] | [ 11,300 ] | [ 5,517 ] |
| Dep. Var. Mean | 0.000 | 0.000 | 0.000 | |
| Productivity Measures: | Teacher Evaluation Overall | Teacher Evaluation Portfolio | Wages in Public Schools | Wages in Private Schools |
| PSU Score | 0.615 *** | 0.477 *** | 0.536 *** | 0.628 *** |
| | ( 0.041 ) | ( 0.04 ) | ( 0.046 ) | ( 0.043 ) |
| (PSU Score)$^2$ | -0.048 *** | -0.031 *** | -0.049 *** | -0.055 *** |
| S | ( 0.001 ) | ( 0.001 ) | ( 0.002 ) | ( 0.002 ) |
| Observations | [ 63539 ] | [ 63539 ] | [ 36771 ] | [ 58523 ] |
| Dep. Var. Mean | 0.000 | 0.000 | 0.000 | 0.000 |
| Employment | Years after graduation | | | Value |
| in Schools | 5 Years | 10 Years | 20 Years | Added |
| PSU Score | 0.298*** | 0.260*** | 0.269*** | 0.334*** |
| | (0.044) | (0.044) | ( 0.089 ) | (0.016) |
| (PSU Score)$^2$ | -0.027*** | -0.025*** | -0.024*** | 0.004 |
| | (0.114) | (0.113) | (0.235) | (0.056) |
| Observations | [ 13,201 ] | [ 13,201 ] | [ 13,201 ] | [3,756] |
| Dep. Var. Mean | 0.470 | 0.435 | 0.287 | 0.000 |

Note: Robust standard errors in parentheses. Significance levels: $*p < 0.10, **p < 0.05, ***p < 0.01$. The regression table reports the coefficients of 14 separate regressions for different teacher performance outcomes over PSU score. Dependent variable PSU is expressed in terms of standard deviations in all cases. In the 'Graduation Time' panel regressions are controlled by year fixed effects and teacher specialization level (primary and secondary levels) fixed effects. In the 'Exit Exam' panel, the dependent variables are % of correct answers except in the writing exam which is a standardized variable; variables are controlled by year and teacher specialization fixed effects. In the 'Productivity Measures' panel, all dependent variables are standardized and regressions are controlled by year and teacher specialization level fixed effects and a second order polynomial for experience. Finally, in 'Works as a Teacher' panel, regressions are controlled by year and teacher specialization fixed effects. In all cases coefficients should be read as the change in the outcome given one standard deviation (SD) of increase in the PSU scores.

# 4 A Carrot and Sticks Approach to Recruiting and Screening

This section studies the implementation of a policy called *Beca Vocacion Profesor* (BVP) which started in 2010. The policy was an aggressive recruiting effort aimed at bringing more qualified students into teacher training programs while limiting access to students with lower pre-college academic achievement. This policy gave full scholarships and other incentives such as stipends and paid semesters abroad for students who matriculate at teaching colleges with scores from approximately the highest 30% of the admissions test distribution. In addition to incentives to individual students, the policy also set incentives for colleges to screen out low performing students. For colleges to be eligible for the scholarship, teacher colleges needed to implement a cutoff score of the 50th percentile of the average score distribution[5]. The results of this "carrots and sticks" policy was that the proportion of high achieving students rose by approximately 50%, although enrollment from the lower end of the distribution continued high at non participating institutions.

## 4.1 BVP Policy Specifics

Specific subsidies for students were determined exclusively by the average math and language college entrance exam score (PSU test). Students who scored above $\mu + \sigma$ (600 points, top 25%) were eligible for a full tuition scholarship. If they scored above $\mu + 2\sigma$ (700 points, top 5%), they were eligible for a full tuition scholarship and a monthly stipend of approx $US150 ( approx 50% minimum wage ). If the student scored above $\mu + 2.2\sigma$ (720 points, top 2%), in addition to a full tuition scholarship and a monthly stipend, they were eligible for a paid semester abroad at a prestigious teaching college. Advertisements mentioned a semester abroad at Stanford or in Finland to name a few.

Requirements for students included having applied and been admitted to an eligible teaching college as a new first year student in 2011 with an entrance exam score from December of 2010. Students previously enrolled in teaching careers are not eligible for this particular modality of the scholarship. The student must have obtained at least $\mu + \sigma$ (600 points) on the entrance exam when calculating the simple average over math and language PSU tests[6]. Beneficiaries have to teach at public schools for three years upon graduation.

Requirements for colleges included being an accredited program for at least 2 years at all campuses. This is determined by the National Commission of Accreditation (CNA). The teaching college must also maintain a minimum score of 500 (the average score) with no more than 15% exceptions among the entering class starting in 2011.

## 4.2 Aggregate Effects on College Participation Choices

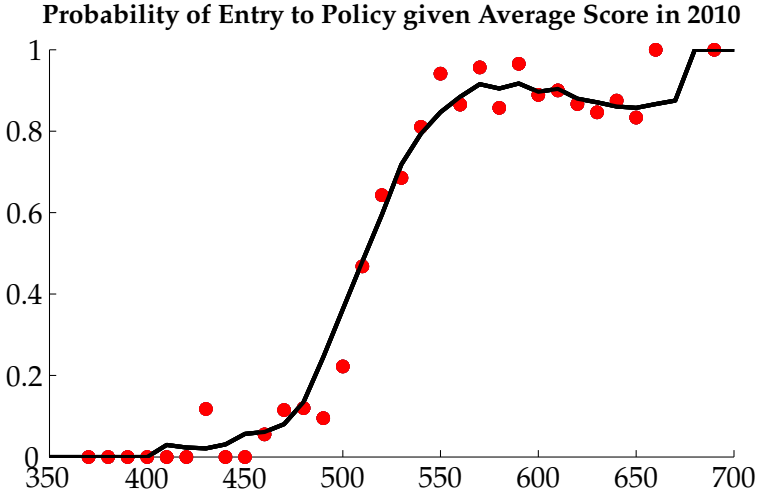
College participation is a key feature of this policy and its effectiveness. The policy offers scholarships to students which is beneficial to colleges competing for students. It also allows colleges to add eligibility to their marketing strategy which can be a signal of quality given it is costly to participate without having a high ability student body. The result was that participation was high among public institutions (forced in fact) and traditional private colleges but participation was very low among institutions with less qualified students and Professional Institutes that also offer teaching degrees. Eligible institutions and careers covered only 40% of matriculated students in 2010. Approximately 1/3 career/college combinations that were eligible did not participate and

[5]The colleges were allowed to accepting up to 15% of their matriculation below that cutoff

[6]If the student had obtained another scholarship called *Beca Excellencia Academica* the cutoff will be 580. These are very few and we ignore students with this scholarship in the analysis.

approximately 1/4 career/college combinations were not even eligible. In general it is safe to assume the screening aspect of the policy had less impactful than it could have been due to low participation of teachers colleges, where only 50% decided to join the program. Figure Figure 7 shows the estimate of participation probability for eligible teacher colleges conditional on the average scores of their freshman class the year prior. Public institutions were forced to participate leading to a positive probability even for very low scores by the figure clearly shows participation being driven by the chosen policy cutoffs.

**Figure 7:** Teaching Colleges Program Participation




Notes: Figure 7 plots the probability of teacher college participation in the BVP policy in 2011 as a function of the average entrance exam of the freshman matriculated there in 2010.

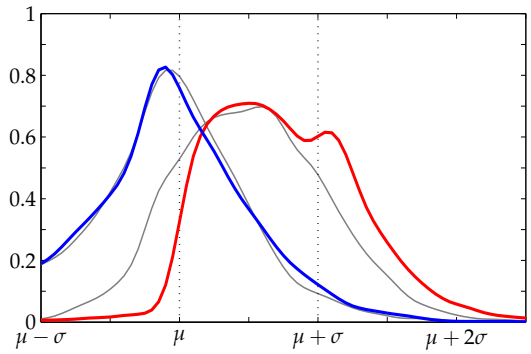
### 4.3 Aggregate Effects on Students Choices

In the application process to the 2011 BVP, the number of applicants to the teacher scholarship was 28,179. Among these, only 3,385 (12%) were eligible for the benefit and 3,252 accepted it (96% of the ones that received it). Among the ones that accepted the fellowship, 3,063 are beneficiaries of the type I fellowship[7], which is the focus of this paper. The effect the policy had on the distribution of scores is quite striking. Figure 8 shows the distribution of test scores for students matriculated in teaching careers and non teaching careers in 2009-2010 and compares the same distributions with 2011 when the policy was in place. Notice the large shift in the density after 600 points which is the first cutoff. Also notice the drop in density below 500 which is due to the restriction on colleges participating (that do not admit students below this threshold). The left panel of Figure 8 shows how the proportion of incoming teacher college students changed at specific intervals of interest.

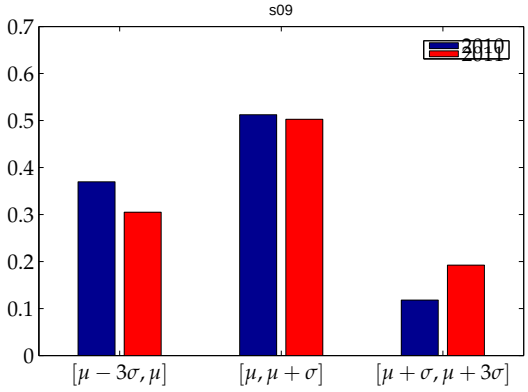
[7]The government implemented two types of fellowships, one for the students that took the standardized test the year before and the other for students that were pursuing careers related to teaching careers and wanted to be teachers. Less than 5% of the winners of the fellowship were applicants through the second type.

**Figure 8:** Aggregate Effects of *Beca Vocacion Profesor*

**(a)** Distribution of Entrance Exam Scores




**(b)** Porportions of Entrance Exam Scores




Note: This figure shows how the composition of students entering teachers colleges changes from 2010 to 2011 after the BVP policy was implemented. Source: MINEDUC and DEMRE.

**??** shows the distribution of students that accepted the BVP by type of enrolled university and PSU score. A 74% were enrolled in universities of CRUCH and 26% in eligible private universities. It can be seen that a disproportionate amount of scholarships go to private school students. However the lions share of the BVP scholarships are going to public and voucher school students.

**Table 2:** Student Participation Shares

| Type of School | Mat in Teaching | Mat at Eligible | Mat with BVP |
|---|---|---|---|
| Public | 41,8% | 38,7% | 34,1% |
| Voucher | 52,6% | 54,2% | 52,9% |
| Private | 5,6% | 7,1% | 13,0% |
| Missing Data | 1,0% | 0,6% | 0,6% |
| Total | 100% | 100% | 100% |

Source: MINEDUC and DEMRE.

## 4.4 Effects on Individual Students Choices

Leveraging the large amount of data, we first plot the conditional choice probabilities and then turn to a formal analysis using a regression discontinuity design. Using a regression discontinuity empirical design we show that the probability of choosing to matriculate in a teaching college rose by 40% to 200% at different point of the distribution. The effects were found to be greater for students who were male and who came from public schools or families with lower income levels.

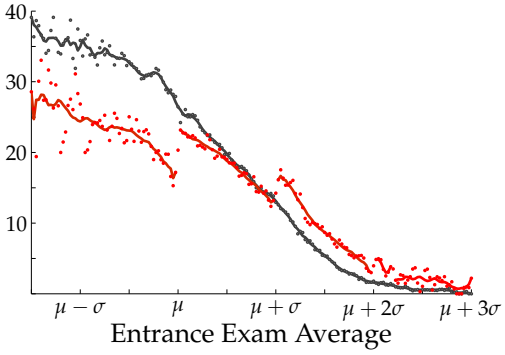
### *4.4.1 Non-parametric analysis*

It is possible to explore empirically the described probability without the necessity of parametric assumptions due to the richness of the data that we have. This will be a non-parametric analysis that will try to describe the probability of choosing a teaching career as a function of the PSU score, eligibility of the fellowship and some demographic characteristics. We estimate the probability as

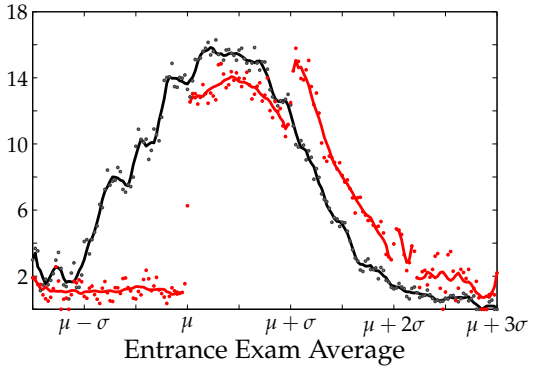
a proportion of students that chose a teaching career between a group of students with similar PSU score. We present in Appendix A the same estimation conditional on the type of school. If we analyze all the teaching careers in the Educational System, we can see a decrease in the probability of studying a teaching career under 500 points and also an increase above 600 points, reaching the maximum level in that score, to then decrease when the PSU score increases.

**Figure 9:** $Pr(\text{Enrollment}|\text{Exam})$ at Teaching Colleges

**(a)** $Pr(\text{Enrollment}|\text{Exam})$ in Any Teaching College



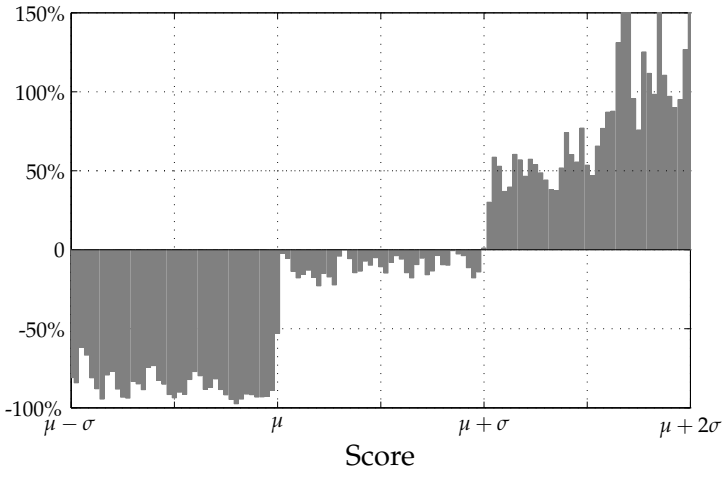

**(b)** $Pr(\text{Enrollment}|\text{Exam})$ in Eligble Teaching Colleges




Note: This figure shows how the conditional probability of enrolling in teaching colleges changed after the introduction of *Beca Vocacion Profesor*. The left panel shows the probability of enrolling in any teaching college and the right panel shows the same but only for programs that opted into the BVP program and were willing to exclude low performing students. Source: MINEDUC and DEMRE.

If we analyze only the probability of choosing an eligible teaching career, the probability for 2011 decreases under the average $\mu$ (500 point average of math and language) almost to zero. To approximate the impact of the fellowship using only the temporal variation, we calculate the change in the probability of choosing a teaching career in several parts of the distribution. Figure 10 shows that there exists an increase close to 40% in the probability of enrollment to a teaching career around $\mu + \sigma$ (600 points), which increases above 100% at $\mu + 2\sigma$ (700 points). Under $\mu$ (500 points) we observe a decrease of 25%. The effect is much larger under the 500 points if we consider only the eligible teaching careers as ca be seen in the right panel of Figure 9. In this case, the probability of studying these careers increase in 40-50% when 600 points is reached, then increase to more than 200% when 700 points is reached. Under 500 points we can observe the probability decrease in almost 100%, which is explain by the restriction of cutoff of 500 points for the careers.

**Figure 10:** Change in $Pr(\text{Enrollment}|\text{Exam})$ in Eligble Teaching Colleges




Note: Figure Source: MINEDUC and DEMRE.

To explore the hypothesis of heterogeneous effects of the fellowship, we observe how the enrollment probability for a teaching career changes for students that came from different types of schools.

$$Pr(E_d|PSU_i = 600, X_i) = Pr(V_(i, E_d) > V_(i, j) forall j available|SchoolType_i, PSU_i)$$

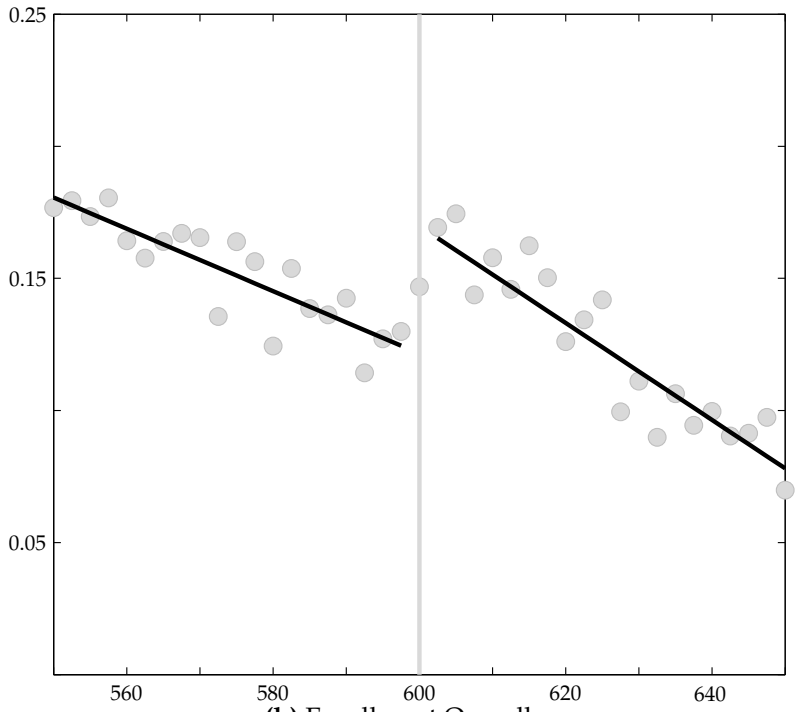
We found a similar pattern in each case. However, the levels are different because the student that came from voucher schools has a higher probability of choosing a teaching career than the students from a private school. Moreover, this difference is even bigger when we compare it with public schools. In terms of the effects of the tuition scholarship, we observe that in all the cases the probability increases if the student has more than 600 points and decreases if the student gets less than 500 points.

### 4.5 Regression Discontinuity Design

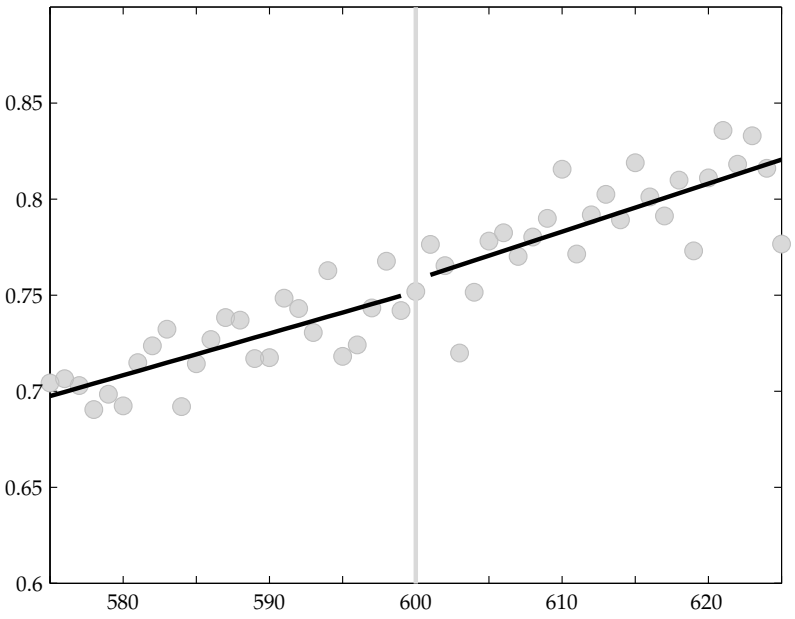
The program in their eligibility criterion uses a discontinuity that could be evaluated using this quasi-experimental methodology. The eligibility criterion is a decision rule that fix a cutoff point to select beneficiaries and non-beneficiaries for the program. This analysis allows estimating the impact of the fellowship around a cutoff point. This is more informative that the non-parametric analysis, but still has some problems when interpreting the results because the analyzed object is a continuous non-parametric function. We will study the effect of the fellowship in terms of the change in the probability $\Delta Pr$ around 600, 700 and 720 points. This methodology does not use many assumptions, but is valid only locally around 600, 700 and 720 points respectively. The intuition behind this method, although it is possible that exists a lot of unobserved heterogeneity in the way the student takes their decision, is the fact that students with similar scores are on average equals. This is, the cutoff point could be approximated to a random neighborhood, in other words, the students that are just above the cutoff point are similar to the student that are just below the cutoff point which requires that both populations are close enough of the cutoff point to secure comparability. The intuition is that students above and below these cutoffs are on average identical so that the noise associated with the test score process effectively give us a local experiment.

Graphically we observe the effect around the 600 points cutoff in Figure 11a. The points are the average among students in groups of 2 points of difference and the line to each side of 600 is a linear regression. Visually, we observe the increase in 600, which will be approximately a jump of 12% to 16%.

**(a)** Enrollment at Teacher Colleges




**(b)** Enrollment Overall




Note: This figure plots the probability of enrollment as a function of college entrance exam test scores.

Specifically, as the next table shown, the increase in the probability in 600 is of 4.2 percentage points on a base of 12, which implies an increase on the percentage close to 35%. Around 700

points we observe an impact of 2.7 percentage points on a base of 2.8. This is equivalent to duplicate the probability of choosing a teaching career. If we consider only the eligible careers, the result is the same, which explains why of this margin, most of the enrollees are in the eligible careers. Finally, when we look at the third level of benefits, around 720 points in the PSU, we could not find evidence of an additional positive effect on the probability of enrollment in a teaching career.

**Table 3:** RDD regression at 600

| | Coefficient | t-stat |
|---|---|---|
| Above 600 | 0.0417*** | 4.00 |
| constant | 0.1204*** | 20.01 |
| (PSU-600) | -0.0007*** | -1.54 |
| | | |
| N=18007 | | |

**Table 4:** RD regression at 700

| | Coefficient | t-stat |
|---|---|---|
| Above 700 | 0.0262*** | 2.42 |
| constant | 0.0283*** | 4.60 |
| (PSU-700) | -0.0015*** | -3.18 |
| | | |
| N=5450 | | |

**Table 5:** RDD regression at 720

| | Coefficient | t-stat |
|---|---|---|
| Above 720 | -0.007 | -0.72 |
| constant | 0.0033*** | 5.38 |
| (PSU-720) | -0.005*** | -1.12 |
| | | |
| N=4150 | | |

#### *4.5.1 Persistent Effects*

Seven years later it can be seen that this policy has had lasting effects. Early productivity indicators such as graduation rates, exit exams and employment probabilities suggest that the policy boosted positive outcomes and raised the quality of students who entered into the teaching profession. In addition, these medium run effects validate the prediction models given that the relationship between pre-college academic Achievement and medium run outcomes is invariant to the recruiting

and screening policies.

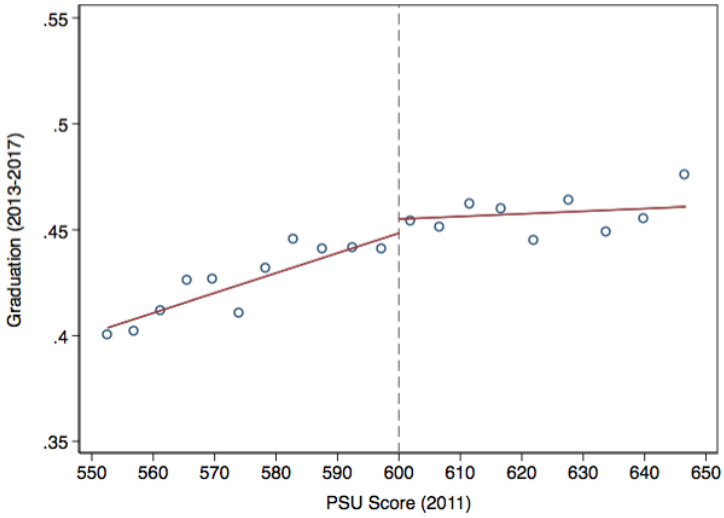
*RDD Estimates on Later Outcomes: Dropout, Graduation, Employment, Exit Exams*

**Table 6:** RDD Estimates for the BVP Scholarship

| | (1) Enrollment in Education | (2) Graduation | (3) Dropout | (4) Employment in Schools | (5) Takes Exit Exam | (6) Score Exit Exam |
|---|---|---|---|---|---|---|
| $\widehat{\alpha_1}$ | 0.052*** | 0.002 | -0.001 | 0.026*** | -0.002 | 0.076 |
| | (0.008) | (0.010) | (0.005) | (0.005) | (0.003) | (0.103) |
| Mean below cutoff | 0.154 | 0.425 | 0.093 | 0.076 | 0.025 | 0.259 |
| Effect Size $(\alpha_1/Mean)$ | 0.339 | 0.004 | 0.008 | 0.343 | 0.063 | 0.295 |
| Observations | 44,418 | 44,418 | 44,418 | 44,418 | 44,418 | 1,079 |

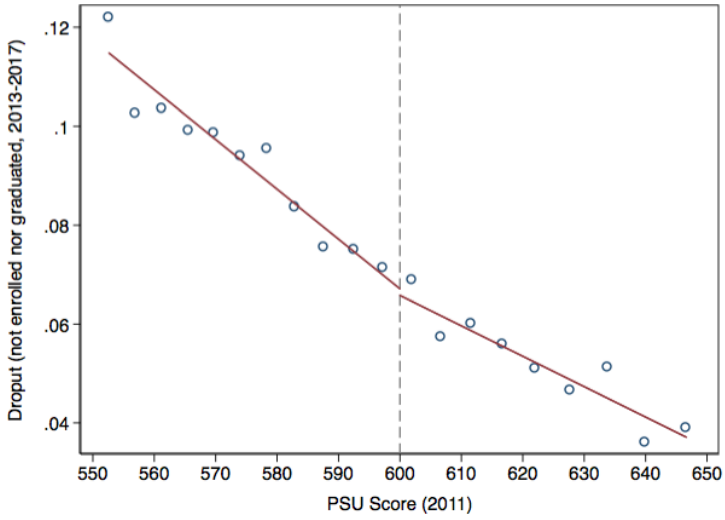
Note: Table 6 shows the results of estimating the equation $Y = \alpha_0 + \alpha_1 BVP + f(Score) + \mu$ where Y are the outcomes in each column, BVP is a dummy for being eligible for the BVB (i.e., having a college entrance exam score of 600 or above, and $f(score)$ is a function of the college entrance exam score, which allows for different slopes on either side of the threshold. The sample size for columns 1 to 5 is of 44,418 students with scores in a window between 550 and 650 points. In column 6 the sample is conditional on taking the exit exam (approximately 2.5% of the sample size of 44,418 students).

**Figure 12:** College Graduation



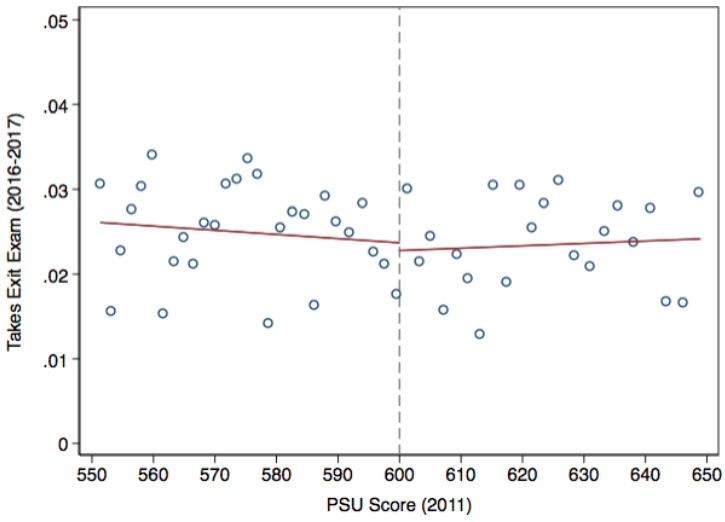

Notes: Figure 12 plots the probability of college graduation between March, 2013 and February, 2018, for all freshmen enrolled in 2011 who took the the college entrance exam in 2011 and where eligible for the BVP scholarship.

**Figure 13:** College Dropout



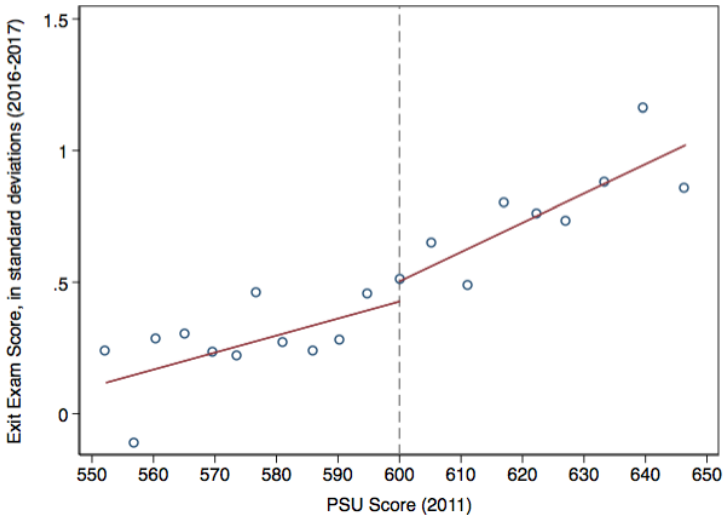

Notes: Figure 13 plots the probability of college dropout, computed as not enrolled nor graduated between March, 2013 and February, 2018, for all freshmen enrolled in 2011 who took the the college entrance exam in 2011 and where eligible for the BVP scholarship.

**Figure 14:** Taking Teacher Exit Exam



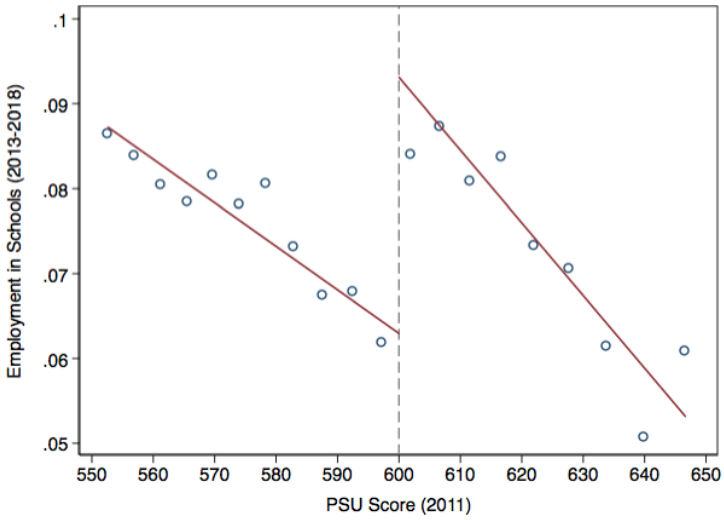

Notes: The figure plots the probability of taking the teacher exit exam in 2016 and 2017, for all freshmen enrolled in 2011 who took the college entrance exam in 2011 and where eligible for the BVP scholarship.

**Figure 15:** Scores in the Teacher Exit Exam




Notes: The figure plots the teacher exit exam score (pedagogical test) in 2016 and 2017, for all freshmen enrolled in 2011 who took the the college entrance exam in 2011 and where eligible for the BVP scholarship.

**Figure 16:** Employment in Schools

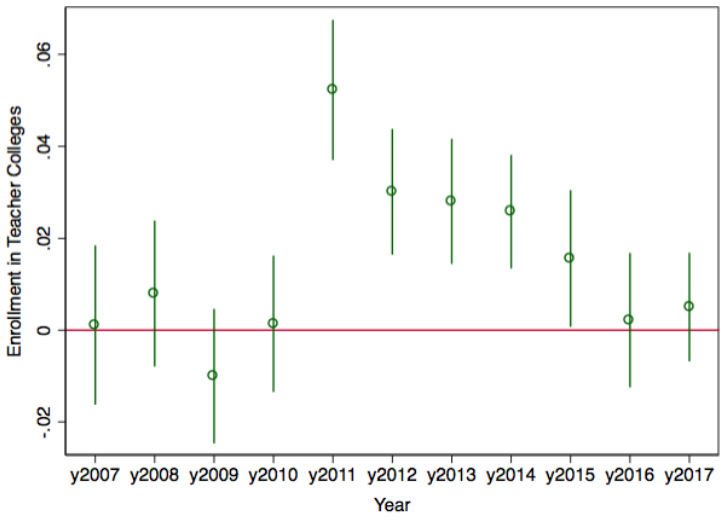



Notes: Figure 16 plots the probability of employment in schools between 2013 and 2018, for all freshmen enrolled in 2011 who took the the college entrance exam in 2011 and where eligible for the BVP scholarship.

*Policy effects depend crucially on other policy*

The analysis above indicates that in 2011 the policy had large effects on individual choice probabilities. However, when viewed over time, the policy effects change dramatically. In particular, in 2016 many colleges in Chile became free of tuition and this would potentially reduce the financial incentive generated by the *Beca Vocacion Profesor* policy. To explore this quantitatively we repeat the RDD exercise for each year and plot the threshold crossing coefficient. Figure 17 shows the evolution of the threshold crossing effect and

**Figure 17:** RDD Estimates on Freshmen Enrollment in Teacher Colleges




Notes: Figure 17 plots the coefficient $\alpha_1$ (with 95% confidence intervals) estimated from the equation $Y = \alpha_0 + \alpha_1 BVP + f(Score) + \mu$ where Y is freshmen enrollment in teacher colleges each year, BVP is a dummy for being eligible for the BVB (i.e., having a college entrance exam score of 600 or above, and $f(score)$ is a function of the college entrance exam score, which allows for different slopes on either side of the threshold. The estimating sample for each year is composed by students with scores in a window between 550 and 650 points.

**Table 7:** RDD Estimates for the BVP Scholarship

| | (1) 2007 | (2) 2008 | (3) 2009 | (4) 2010 | (5) 2011 | (6) 2012 | (7) 2013 | (8) 2014 | (9) 2015 | (10) 2016 | (11) 2017 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\widehat{\alpha_1}$ | 0.001 | 0.008 | -0.010 | 0.001 | 0.052*** | 0.030*** | 0.028*** | 0.026*** | 0.016** | 0.002 | 0.005 |
| | (0.009) | (0.008) | (0.007) | (0.008) | (0.008) | (0.007) | (0.007) | (0.006) | (0.008) | (0.007) | (0.006) |
| Mean below cutoff | 0.165 | 0.157 | 0.161 | 0.161 | 0.154 | 0.141 | 0.118 | 0.113 | 0.119 | 0.114 | 0.119 |
| Effect Size ($\alpha_1 / Mean$) | 0.007 | 0.051 | 0.062 | 0.008 | 0.339 | 0.214 | 0.238 | 0.228 | 0.131 | 0.019 | 0.042 |
| Observations | 36,077 | 37,282 | 41,208 | 43,954 | 44,418 | 44,247 | 44,251 | 44,056 | 46,419 | 44,191 | 44,975 |

Note: Table 7 shows the results of estimating the equation $Y = \alpha_0 + \alpha_1 BVP + f(Score) + \mu$ where Y is freshmen enrollment in teacher colleges each year in each column, BVP is a dummy for being eligible for the BVB (i.e., having a college entrance exam score of 600 or above, and $f(score)$ is a function of the college entrance exam score, which allows for different slopes on either side of the threshold. The estimating sample for each year is composed by students with scores in a window between 550 and 650 points.

### *4.5.2 Discussion of BVP Policy*

At the aggregate level, the policy increases the average quality of predicted teacher effectiveness. Overall, it is shown that the majority of the effects on the average student entering teaching college were due to the incentives to screen out low performing students and not the recruitment effects on their own. The reason is that while free tuition and additional subsidies raised choice probabilities significantly, the number of students at those margins is small relative to the population of students in teaching colleges. It is estimated a total of 1000 additional students entered teaching colleges from the top 30% of the distribution but the total number of freshman students at teachers' colleges is closer to 20000. Screening restrictions are estimated to have reduced the bottom tail of the distribution by 4000. In terms of predicted productivity, the concavity of the mapping from pre-college academic Achievement and predicted teacher effectiveness also imply that reducing low performers is has more impact than adding very high performers.

## 5 A Mandatory Screening Policy

This section studies a mandatory screening policy implemented in 2017. The Chilean Government enacted the *NUEVA LEY DE CARRERA DOCENTE* (NLCD)[8]. This is a broad policy that creates a new system of professional development for teachers in the country. The Law includes specific guidelines for the recruitment, development and retention of teachers. One important aspect of the policy is that only accredited programs can now confer teaching degrees. To be accredited, all teacher colleges are required to implement an entrance exam and the exit exam within the last twelve months before students graduate. To be accredited colleges must also comply with new admissions screening rules based on pre-college academic achivement, specifically making use of college entrance exams and high school GPA rank within school.

In this section we explore the consequences of the screening policies in two distint ways. The first is to apply the screening policy to prior cohorts and determine what the partial equilibrium effects would have been in terms of total matriculation and also to categorize the students that were screened out or in by their ex-post outcomes. In this excersize we expect the screenig rule to be successful at blocking entry to teaching colleges to students who did not graduate, did not

[8]The law 20.903 is available here Law 20,903

get teaching jobs, or went on to become less effective teachers. We would expect a successful screening rule to also not leave out students who went on to be highly effective teachers as well.

In a second way to explore the effects of the screening policy is to use the best simple prediction model above to evaluate how the policy affected predicted teacher outcomes such as graduation, exit exams, getting a job, becoming an effective teacher. We present this calculation for each of the last ten years to put the magnitude of the screening policy into perspective. Finally, in a third excersize, we use flexible prediction models to search among a family of simple but flexible screening rules that can provide the same screening power rejecting the same amount of low performing teachers while minimizes the mistakes made when rejecting future effective teachers.

### 5.1 NLCD Policy Specifics

The requirements for the screening policy affects admissions to all teacher colleges and are designed to be implemented gradually. During the first three years (2017-2019), the screening policy requires students to either haveachieved an entrance exam score above the 50th percentile of the distribution when averaging math and language. Alternatively, students can also avoid the screening rule if their high school GPA is above the 70th percentile within their high school in their graduating cohort. For the admissions cycle of 2020 to 2022, teh screening rule increases the requirements. Students must have achieved an exntrance exam score above the 525 points when averaging math and language scores or have GPA above the 80th percentile. In addition, if students have a GPA above the 60th percentile and test scores in math and language that average above 500 points, they may also matriculate in teacher colleges. Finally, in 2023 and onwards, the screening policy reaches its steady state and requires students to have entrnce exam scores above 550 points (the 70th percentile) or be in the top 10% of their cohort GPA. If the student is in the top 30% of the GPA distribution at their high school **and** has scores above the average, then that student can also matriculate at teaching colleges.

All of these conditions are designed as minimal requirements for admission to teacher colleges. Each institution is allowed to consider stricter conditions, define number of vacancies or slots and application mechanisms. However, all the requirements must be informed before the beginning of admission process each year.

### 5.2 Aggregate Effects on Teacher College Matriculation

[*place figures of distribution and porportions as above in BVP*]

## 5.3 Simulating the Screening Rule Back In Time

**Figure 18:** Screening Rule Applied To Past Teacher College Students

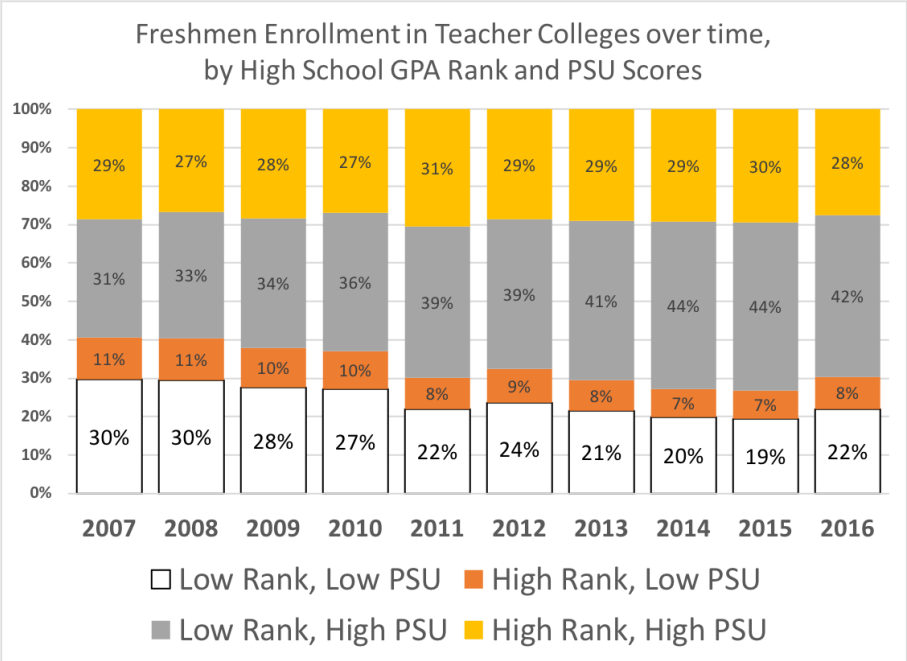
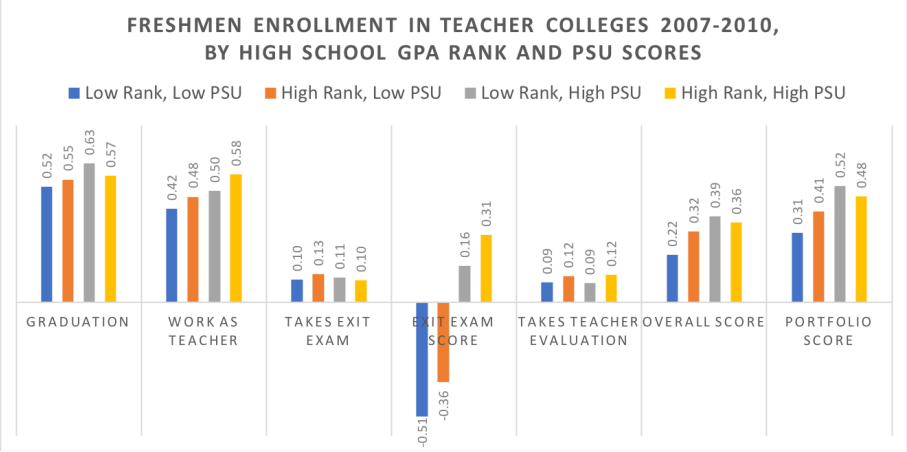


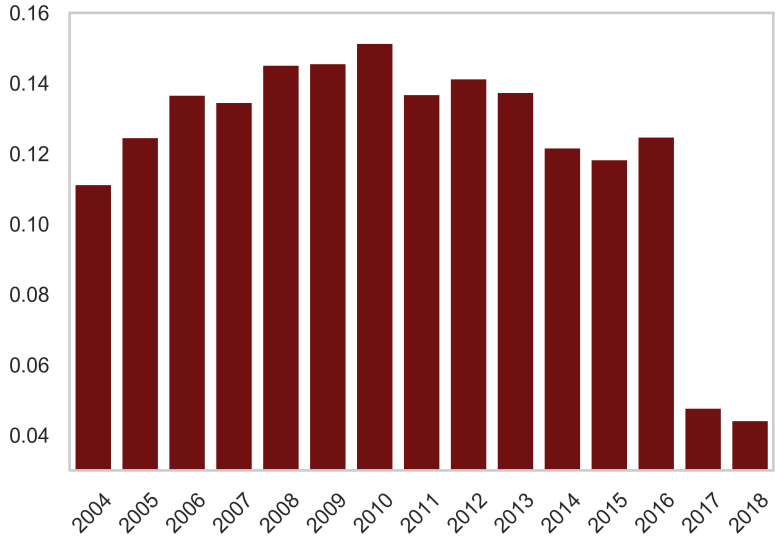

**Figure 19:** Outcomes for Those Screened In Simulation




## 5.4 Simulating the Screening Rule Using Predicted Outcomes Going Forward

The MINEDUC rule for screening consists of allowing prospective students to enter a program in education if their average score in mathematics and verbal is higher than 500 or they belong to the 70th percentile in high school. We use the MINEDUC rule to identify teachers that would have been characterized as bad performers in the whole sample of teachers until 2016 and test how much would a Machine Learning rule can improve the stock of selected teachers guaranteing that we equate the same number of accurately predicted teachers as either good or bad from the whole sample.

**Figure 20:** Percentage of Predicted Share of Unsatisfactory Teachers

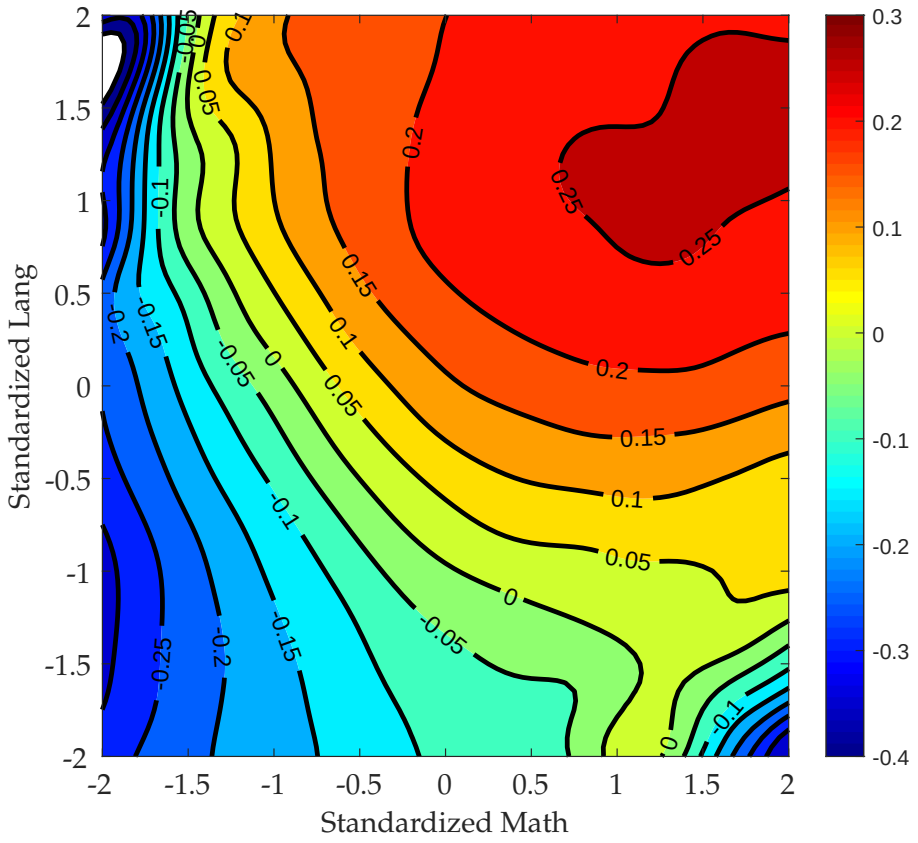



## 5.5 Towards A More Efficient Screening Policy

Increased availability of data and algorithms can help policy-makers make better predictions. In this case, a better prediction model can help guide the design of a more efficient data driven screening policy. Building on evidence presented above, we now try to improve on the simulated outcomes produced by the current policy under NLCD. In this excercise, our predictive features are scores on the entrance exam (all of them) and the relative ranking of students in their last year of high school as well as their relative ranking in math, verbal, art, science, and social sciences. Our main outcome variable takes the value of one (or positive) if the teacher was qualified as unsatisfactory by the goverment (or being below the 30 percentile), and zero (or false) if not. For this, we start by showing the predictive performance of the government rule (GR).

To highlight the potential role for a flexible screening rule, Figure 29 shows the conditional mean off teacher evaluations conditional on math and language entrance exam scores. The simple average of math and language scores would cut across this space in a linear way but it can be seen that level curves are quite nonlinear. This suggests a modification of the rule that puts equal weights on math and language at an arbitrary cutoff is likely going to have less success than a more flexible rule.

**Figure 21:** Contour Plot of the Estimated Conditional Mean of Observational Teacher Evaluations




Note: This figure shows the contour plot of the mean of teacher evaluations conditional on pre-college academic achivement in math and language scores. See **??** for a summary of the correlation of pre-college academic achivement and other measures of teacher productivity.

The steps of our analysis are as follows:

1. Divide the sample in training and validation set.
2. Compute a grid search of different models with different hyperparameters over the training data.
3. Crossvalidate results over each model in the grid for the training data which will be split into training and test datasets.
4. Compare the teacher evaluation score distribution of teachers accepted by the new rule and the NLCD rule using the validation data.

We implement a series of simple flexible prediction models and we present the results below. A tabulation of the GR in our validation data suggests that 83% of teachers would have been left out by the MINEDUC rule if they applied under the GR scheeme. From the total of teachers predicted as unsatisfactory, 68% were actually qualified as satisfactory by the government and 31% as unsatisfactory. From the 17% of teachers predicted as satisfactory, 81% of them were reported as satisfactory and the rest 19% as unsatisfactory.

**Table 8:** Performance Measures given same Accuracy Level as Government Rule

| | ROC-AUC | Cutoff | Accuracy | Recall | Precision | F1-Score |
|---|---|---|---|---|---|---|
| KNearestRegressor | 0.610 | 0.025 | 0.400 | 0.880 | 0.316 | 0.736 |
| BaggingRegressor | 0.609 | 0.024 | 0.400 | 0.879 | 0.316 | 0.736 |
| LogisticRegression | 0.612 | 0.781 | 0.400 | 0.877 | 0.315 | 0.735 |
| OLS | 0.614 | 0.025 | 0.400 | 0.876 | 0.315 | 0.735 |
| Lasso | 0.614 | 0.025 | 0.400 | 0.876 | 0.315 | 0.735 |
| RandomForestRegressor | 0.608 | 0.025 | 0.400 | 0.857 | 0.313 | 0.733 |
| KNearest | 0.607 | 0.778 | 0.396 | 0.877 | 0.314 | 0.736 |
| AdaBoostRegressor | 0.574 | 0.024 | 0.385 | 0.876 | 0.310 | 0.739 |
| RandomForest | 0.610 | 0.775 | 0.400 | 0.862 | 0.313 | 0.733 |
| Adaboost | 0.594 | 0.506 | 0.399 | 0.851 | 0.312 | 0.732 |
| GovernmentRule | nan | 5.000 | 0.400 | 0.868 | 0.314 | 0.734 |

Note: We applied grid search over the hyperparameter space of each model and crosvalidated their results over different test samples for each of them to find the best specification. The main metric to select the best specification was the ROC-AUC for classification models and R2 for regression models. All measures of Accuracy, Recall, Precision and F1-Score were computed based on prediction results of fixing a cutoff value that equalizes the accuracy level of the Government Rule.

Table 8 summarizes our best models for each of the algorithms tested[9]. The cutoff column represents the cutoff that guarantees a equivalent accuracy measure to the government rule, as seen in the accuracy column. The area under the curve shows results ranging from 0.60 to 0.62 for all the models. The graphical representation of the AUC can be seen from Figure 22.

We tested all the ML recruitment policies with the reported cutoffs in table 8 and identified good and bad performing teachers. After it, we compared the portfolio score distribution of teachers selected by the different ML algorithms versus the ones selected by the GR in our validation data. Table 9 suggests that using any of the ML methods contributes positively to the score distribution of accepted teachers relative to those accepted by the government rule. The best one appears to be the KNearest Regressor which increases the average portfolio score of accepted teachers in about 6% of a standard deviation, and in second place the Random Forest increasing 5% of a standard deviation.

Our best model KNearest Regressor, by construction, based on a cutoff equivalent to the GR accuracy performance reports similar percentages for false and true positives or negatives. However, their mistakes are less serious. As can be seen from Table 9 truly predicted good performers are performing 4% SD better on average than those from the GR. Also, from those accepted and erroneously predicted as good performers their scores are on average 2.5% of a SD higher than the GR. Also among the rejected ones, those that were erroneously predicted as bad performers, their average score was less higher by 1% of a SD. All these numbers combined contribute to a 6% SD increase for the average score of accepted teachers and 1% SD reduction in the score of non selected teachers. Figure 24 shows how the score distributions of each group change; they skew more to the right in the case of true negatives and false negatives, which means we are recruiting

[9]We ran classification models such as Logitstic Regression, Random Forest, K-Nearest, and Adaboost, and regression models such as AdaBoost Regressor, Random Forest Regressor, KNearest Regressor, OLS, Lasso, Bagging Regressor. The classification models estimated a latent variable out of the dichotomous variable computed out of the portfolio score, while the regression models regressed the predictors against portfolio score.

**Table 9:** Policy Rule Contribution to Recruited Teacher Performance

| | Accepted | Rejected | True P. | True N. | False P. | False N. |
|---|---|---|---|---|---|---|
| KNearestRegressor | 6.508 | -1.111 | 0.228 | 5.823 | -1.336 | -0.798 |
| BaggingRegressor | 5.813 | -1.004 | 0.003 | 4.897 | -1.125 | 0.740 |
| LogisticRegression | 5.670 | -1.014 | 0.100 | 5.263 | -1.233 | -0.069 |
| OLS | 5.004 | -0.912 | 0.597 | 5.703 | -1.360 | -3.678 |
| Lasso | 5.004 | -0.912 | 0.597 | 5.703 | -1.360 | -3.678 |
| RandomForestRegressor | 4.258 | -1.222 | 0.447 | 8.583 | -2.330 | -3.337 |
| KNearest | 3.661 | -0.415 | 0.405 | 4.394 | -0.869 | -2.252 |
| AdaBoostRegressor | 3.456 | 0.032 | -0.550 | 5.805 | -0.806 | 4.407 |
| RandomForest | 1.670 | -0.514 | 0.137 | 3.720 | -1.010 | -1.240 |
| Adaboost | -0.073 | -0.320 | 0.112 | 3.935 | -1.148 | -1.615 |

Note: Models resulted from Table 8 were sorted by its contribution to the accepted teachers' performance score. First we fixed cutoff levels for each model to the Government Rule accuracy level and computed the mean differences on teaching performance for the individuals in different groups. The (1) Accepted ones which are those that were predicted as good performers, the (2) Rejected or predicted as bad performers, the (3) True Negatives or accurately predicted as good performers (4) The true Positives or accuractely predicted as bad performers (5) False Positives or erroneously predicted as bad performers, (6) and the False Negatives or erroneously predicted as good performers. The values in the table show the change on portfolio score as a percentage of a standard deviation in the test.

more qualified teachers.

Figure 26 shows what contributes the most to the accuracy of the model. As the number of observations increases, the accuracy of the model increases but has a concave shape. The increase occurs until we reach the 5000 observations. After that number the observations, the contribution is centered around 61,6%, however it reduces the variance which means having less noisy cross-validated measures of performance. Variables seem to be contributing the most to our prediction performance, for instance, scores in Verbal reduce the R2 in the validation data in about 15% when removed from the estimation and 11% in the case of Math scores as seen from figure 26.

### 5.6 Discussion of Empirical Results

In the previous sections we have found evidence that students with low levels of pre-college academic achievement tend to be less effective teachers but high scoring students, while likely to be above average teachers, are similar in expectation to above average students. These correlations imply that pre-college academic achievement seems to be a necessary but not sufficient condition to be a great teacher and a not having a minimal level of preparation is predictive of being a less effective teacher later on.

The recruiting and screening policies presented suggest that incentves matter in bringing talent to the teacher profession but if targeted at very high scoring students, the aggregate effects are minimal due to strong outside options. When new government policies expanded free college to other majors, the incentives of the recruiting policy became ineffective.

Overall the main driving force moving the average teacher predicted quality seemes to have been screening policies that limit entry into the teaching profession. We show that these policies can be made more effective if the rules governing them are data driven and more flexible.

# 6 Conclusions

Data is becoming more abundant as administrative sources becomes available, historical information gets digitalized and new information gets recorded in more detail than ever before. This data, together with improved algorithms for making use of it, is lowering the cost of making increasingly accurate predictions and is influencing decisions, such as hiring, in many markets (Agrawal et al., 2018). In this paper we show evidence that in the context of a developing country, pre-college academic achievement is systematically related to a series of measures of long run teacher productivity. We then evaluate the effectiveness of two recent policies that looked to screen or recruit based on pre-college academic performance and find that both raise the predicted quality of teachers, own graduation rates and exit exams scores. We argue the results indicate that policies that use pre-college academic achievement to either screen out or incentivize an application to teacher colleges can be feasible and useful in some contexts. This is particularly important because teacher labor markets are known to be inefficient and, in many cases, there is limited scope to sideline or retrain ineffective teachers once they are in the system. If teacher effectiveness was possible to predict early on, policies could focus resources on recruiting and retaining the most promising candidates and filtering out applicants who are more likely to be ineffective teachers later on.

In our analysis we show that very simple models can use the data to categorize and identify a significant proportion of the worst performing teachers but simultaneously also fail to identify the very effective ones. We interpret this concave relationship between pre-college academic achievement and later productivity as evidence that in a developing country context such as Chile, basic academic competency is a necessary condition to be an effective teacher. We show suggestive evidence that this relationship between pre-college academic achievement and productivity is seemingly not caused by high scoring students having differential access to more selective and more effective teaching colleges. In fact, we find no meaningful differences across different teaching colleges when using an regression discontinuity strategy build on the centralized assignment mechanism used in Chile since the 1970s. Taken together, our evidence seems to suggest that neither training or experience are enough to undo the significant initial deficiencies that very low performing students have and thus systematically more likely to be low performing teachers thirty and forty years later when observed in the classroom.

We then evaluate two policies implemented in Chile that look to shape the pool of students entering teaching colleges by screening out low performing students or setting incentives for high performing students based on their pre-college academic achievement. We show evidence that financial incentives such as scholarships and stipends do indeed lead to higher recruitment of very talented students entering teacher colleges but the effectiveness depends crucially on the outside options high achieving students have. The first year of the policy, students with over $1\sigma$ average math and language college entrance exam were provided with full tuition when they studied at a participating teaching college. No other similar scholarships were available at that time. Colleges that were eligible had to limit entry to only students above average. Half of the teacher colleges decided to participate significantly lowering the amount of low performing students matriculating in teacher colleges nationwide. The financial incentives attracted more high scoring students and raised the probability a student with $1\sigma$ above the mean would choose to be a teacher by 40% but on a base of 10%. Six years later, students graduated and performed as predicted their higher observables lending credence to the ability for the relationship between pre-college academic ability and later outcomes is invariant to these types of policies. A few years later, many higher education options became tuition free as part of a new government policy years later, the effectiveness of the financial incentives was reduced significantly. The results suggest that induc-

ing colleges to voluntary exclude the lowest performing students was the most effective aspect of the policy. The results also highlight that the effectiveness of targeting highly talented students with recruiting efforts is highly context dependent and expensive because they have many other valuable options.

A second screening policy implemented in 2017 barred teaching colleges from admitting students with below average scores unless they had very high GPA. This policy was found to be effective at shifting the predicted quality of future teachers, screening out between 20-30% of the least academically prepared applicants. To evaluate the policy relevance of a minimum standard for entering teaching colleges, we develop a model that classifies potential teacher productivity based on the rich set of pre-college information including GPA course transcripts and entrance exam scores. This model provides feasible cutoff rules that exclude students with a higher chance of being a low performing teacher. Partial equilibrium analysis shows that if implemented, these rules would have bound for 20% of students entering teaching colleges in 2016 and would have affected 15-25% of current teachers, including 40% of the worst performers based on government teacher evaluations. We interpret these results as suggestive that screening policies can be improved with even simple models and a data driven policy rule.

In both policies studied, the increase in the distribution of predicted teacher productivity comes with screening policies aimed at excluding the very lowest performing students and not with recruiting the highest ability students. This is both a function of the higher ability to identify low productivity teachers from the bottom of the academic achievement distribution and that recruiting high ability students is hard. Taken together this suggests that increasing the predicted productivity of a cohort of future teachers can be increased first by excluding the lower tail of distribution of academic achievement and potentially using any resources saved to incentivize a large group of simply above average students to enroll in teaching colleges with the former being the more effective of the two.

The evidence presented in this paper could be viewed as being at odds with the consensus in the US literature that teacher effectiveness is not very predictable Rothstein (2015). Part of the discrepancy could be due to better data availability in Chile at this time. In our view, while this is likely, the more relevant issue is that the context is very different and the distribution of underlying academic preparation in a country like Chile is shifted to the left and has a wider dispersion relative to the USA. Exit exams in Chile indicate more than 40% of teachers do not pass a subject test on their own material they are supposed to teach. It is quite possible that in a context such as the US, the relationship between pre-college academic achievement and later effectiveness is already on the flat part of **??** and these results are mostly relevant for developing countries.

The policy relevance of screening policies for developing countries can be seen as limited because they are usually trying to expand educational access and need more teachers and more classrooms. The temptation to expand rapidly is to lower standards and fill classrooms with bodies. However, these newly minted teachers with potentially low ability and limited prospects will seemingly continue to be as such for the next thirty or forty years as we found in the data for Chile. In this context it might do a country well to consider growing more slowly, sticking with minimal standards for entry into the teaching profession and higher wages, with a mind to planning ahead towards a smoother transition from a system that provides quantity to one that provides quality. These results are also important for countries that like Chile, have seen a tremendous growth in the supply of higher education options. Teaching is a relatively cheap degree to offer and supply expanded faster than any other option in Chile after government backed loans were provided by the government for the first time. Many students with low scores then find themselves with limited options but teaching is virtually always feasible. Minimal standards for entry or for access to subsidies, can also help regulate the supply of degrees that are being oversupplied by reducing

demand from groups that less likely to benefit from those studies.

In this paper we have outlined that screening and recruiting policies implemented before candidates enter college could be feasible and useful in some context. A data driven approach to determining the specific details of the policies seem promising. Future work should study the equilibrium effects of these policies as they will likely affect incentives for universities. Research is needed to understand how to improve models and data to better screen candidates (or to realize they should not screen) in new context and considering the objectives and priorities of the policy maker. It is important to mention that every empirical model evaluated would have excluded some very effective teachers as does the current screening policy in Chile indicating there is no free lunch. More research is needed to understand how to make policy rules of this type are fair and not systematically biased against subgroups of the population.

## References


**Agrawal, Ajay, Joshua Gans, and Avi Goldfarb**, *Prediction machines: the simple economics of artificial intelligence*, Harvard Business Press, 2018.

**Araujo, M Caridad, Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady**, "Teacher quality and learning outcomes in kindergarten," *The Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

**Bacolod, Marigee**, "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality," Working Papers, U.S. Census Bureau, Center for Economic Studies 2006.

**Bau, Natalie and Jishnu Das**, "Teacher Value-Added in a Low-Income Country," *American Economic Journal: Economic Policy*, 2018.

**Bruns, Barbara and Javier Luque**, *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*, Washington, DC: World Bank, 2015.

**Chaudhury, Nazmul, Jeffrey Hammer, Michael Kremer, Karthik Muralidharan, and F. Halsey Rogers**, "Missing in Action: Teacher and Health Worker Absence in Developing Countries," *Journal of Economic Perspectives*, Winter 2006, *20* (1), 91–116.

**Corcoran, Sean, William Evans, and Robert M. Schwab**, "Changing Labor-Market Opportunities for Women and the Quality of Teachers, 1957-2000," *American Economic Review*, 2004, *94* (2), 230–235.

**Gallegos, Sebastian and Christopher Neilson**, "50 years of College Entrance Exams," *Princeton University, Mimeo*, 2017.

**Gregorio, Soledad De, Sebastian Gallegos, and Christopher Neilson**, "Decomposing Teachers Contributions to Learning," Technical Report.

**Hanushek, Eric and Richard R. Pace**, "Who chooses to teach (and why)?," *Economics of Education Review*, 1995, *14* (2), 101–117.

**Hastings, Justine, Christopher Neilson, and Seth Zimmerman**, "Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile," Working Paper 19241, NBER 2014.

**Hoxby, Caroline and Andrew Leigh**, "Pulled Away or Pushed Out? Explaining the Decline of Teacher Aptitude in the United States," *American Economic Review*, 2004, *94* (2), 236–240.

**Kane, T. and S. Cantrell**, "Learning about teaching research report: Initial findings from the measures of effective teaching project," Technical Report 2010.

**Manski, Charles**, "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence From the National Longitudinal Study of the High School Class of 1972," NBER Working Papers 1539, National Bureau of Economic Research, Inc 1985.

**MINEDUC**, "Sistema de Informacion General de Estudiantes: Idoneidad Docente," Data File, Retrieved from http://datos.mineduc.cl/dashboards/19732/bases-de-datos-de-cargos-docentes/, Santiago, Chile: Ministerio de Educacion de Chile 2019.

**Mizala, Alejandra and Ben Ross Schneider**, "Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990-2010)," *Governance*, 2014, *27* (1), 87–109.

_ **and Hugo Nopo**, "Measuring the relative pay of school teachers in Latin America 1997–2007," *International Journal of Educational Development*, 2016, *47* (C), 20–32.

**Neilson, Christopher A.**, "The Rise of Centralized Mechanisms in Education Markets Around the World," Technical Report August 2019.

**OECD**, *Education at a Glance 2009: OECD Indicators*, Paris: Organisation for Economic Cooperation and Development, 2009.

**Paredes, Valentina, Andreas Aron, and Alvaro Carril**, "Where is the Teacher? Short-run effect of teacher Absenteeism on Student Achievement," Technical Report, Department of Economics, Universidad de Chile. Mimeo. 2015.

**Podgursky, Michael, Ryan Monroe, and Donald Watson**, "The academic quality of public school teachers: an analysis of entry and exit behavior," *Economics of Education Review*, 2004, *23* (5), 507–518.

**Rockoff, Jonah E.**, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review*, 2004, *94* (2), 247–252.

**Rothstein, Jesse**, "Teacher Quality Policy When Supply Matters," *American Economic Review, 105(1): 100-130*, 2015.

**Vegas, Emiliana, Richard Murnane, and John Willet**, "From High School to Teaching: Many Steps, Who Makes It?," *Teachers College Record*, 2001, *103* (3), 427–449.

**World Bank**, *Population aging: Is Latin America Ready?*, Washington, DC: David Cotlear (editor), 2011.

_ , "System Approach for Better Education Results (SABER): What Matters Most in Teacher Policies? A Framework for Building a More Effective Teaching Profession. Washington, DC.," Technical Report 2013.

**Table 10:** Freshmen PSU (Raw) Scores and Enrollment in Teacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 15,057 | 542 | 70 | 456 | 506 | 540 | 585 | 627 |
| 2017 | 13,507 | 542 | 70 | 456 | 508 | 542 | 584 | 627 |
| 2016 | 15,323 | 525 | 77 | 422 | 482 | 528 | 574 | 618 |
| 2015 | 14,453 | 531 | 78 | 428 | 490 | 534 | 582 | 626 |
| 2014 | 13,948 | 532 | 77 | 432 | 492 | 534 | 581 | 626 |
| 2013 | 15,488 | 527 | 76 | 424 | 485 | 529 | 577 | 620 |
| 2012 | 18,454 | 524 | 77 | 420 | 481 | 526 | 574 | 618 |
| 2011 | 19,597 | 528 | 78 | 425 | 484 | 530 | 580 | 622 |
| 2010 | 20,575 | 517 | 72 | 425 | 478 | 518 | 564 | 604 |
| 2009 | 19,226 | 516 | 72 | 423 | 476 | 518 | 564 | 604 |
| 2008 | 18,485 | 510 | 75 | 410 | 468 | 514 | 560 | 602 |
| 2007 | 18,070 | 510 | 76 | 403 | 466 | 516 | 562 | 603 |

**Table 11:** Freshmen PSU Z-Scores and Enrollment in Teacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 15,057.00 | 0.34 | 0.65 | -0.45 | 0.01 | 0.33 | 0.74 | 1.13 |
| 2017 | 13,507.00 | 0.38 | 0.64 | -0.41 | 0.07 | 0.38 | 0.77 | 1.16 |
| 2016 | 15,323.00 | 0.22 | 0.70 | -0.72 | -0.18 | 0.25 | 0.67 | 1.07 |
| 2015 | 14,453.00 | 0.25 | 0.71 | -0.70 | -0.13 | 0.29 | 0.72 | 1.11 |
| 2014 | 13,948.00 | 0.26 | 0.70 | -0.66 | -0.11 | 0.28 | 0.71 | 1.11 |
| 2013 | 15,488.00 | 0.21 | 0.70 | -0.72 | -0.17 | 0.23 | 0.67 | 1.07 |
| 2012 | 18,454.00 | 0.19 | 0.69 | -0.74 | -0.20 | 0.21 | 0.64 | 1.04 |
| 2011 | 19,597.00 | 0.25 | 0.71 | -0.69 | -0.15 | 0.27 | 0.72 | 1.11 |
| 2010 | 20,575.00 | 0.15 | 0.65 | -0.69 | -0.20 | 0.16 | 0.58 | 0.94 |
| 2009 | 19,226.00 | 0.14 | 0.66 | -0.71 | -0.22 | 0.17 | 0.59 | 0.94 |
| 2008 | 18,485.00 | 0.09 | 0.68 | -0.83 | -0.30 | 0.13 | 0.55 | 0.93 |
| 2007 | 18,070.00 | 0.09 | 0.70 | -0.88 | -0.32 | 0.14 | 0.56 | 0.94 |

# Tables

**Table 12:** Freshmen PSU (Raw) Scores and Enrollment in CRUCH Teacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 9,484 | 561 | 64 | 496 | 520 | 557 | 601 | 640 |
| 2017 | 8,579 | 561 | 62 | 502 | 522 | 556 | 599 | 639 |
| 2016 | 7,703 | 563 | 61 | 503 | 524 | 558 | 600 | 641 |
| 2015 | 7,360 | 571 | 59 | 509 | 530 | 565 | 608 | 646 |
| 2014 | 6,800 | 570 | 59 | 508 | 528 | 564 | 608 | 648 |
| 2013 | 7,233 | 566 | 60 | 506 | 526 | 560 | 604 | 643 |
| 2012 | 7,264 | 571 | 55 | 510 | 530 | 564 | 605 | 643 |
| 2011 | 7,837 | 576 | 56 | 513 | 536 | 570 | 612 | 648 |
| 2010 | 8,539 | 558 | 56 | 492 | 520 | 557 | 595 | 629 |
| 2009 | 8,403 | 557 | 55 | 492 | 520 | 556 | 593 | 626 |
| 2008 | 8,190 | 554 | 57 | 488 | 517 | 553 | 592 | 625 |
| 2007 | 8,621 | 549 | 64 | 478 | 513 | 552 | 590 | 623 |

**Table 13:** Freshmen PSU Z-Scores and Enrollment in CRUCH Teacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 9,484.00 | 0.51 | 0.59 | -0.10 | 0.14 | 0.48 | 0.89 | 1.25 |
| 2017 | 8,579.00 | 0.55 | 0.57 | 0.01 | 0.20 | 0.51 | 0.90 | 1.27 |
| 2016 | 7,703.00 | 0.56 | 0.56 | 0.02 | 0.22 | 0.53 | 0.91 | 1.28 |
| 2015 | 7,360.00 | 0.61 | 0.53 | 0.05 | 0.25 | 0.56 | 0.95 | 1.30 |
| 2014 | 6,800.00 | 0.61 | 0.54 | 0.04 | 0.23 | 0.55 | 0.96 | 1.32 |
| 2013 | 7,233.00 | 0.57 | 0.55 | 0.02 | 0.20 | 0.51 | 0.92 | 1.27 |
| 2012 | 7,264.00 | 0.61 | 0.50 | 0.06 | 0.24 | 0.55 | 0.92 | 1.26 |
| 2011 | 7,837.00 | 0.68 | 0.51 | 0.11 | 0.31 | 0.64 | 1.01 | 1.34 |
| 2010 | 8,539.00 | 0.52 | 0.51 | -0.09 | 0.17 | 0.51 | 0.86 | 1.17 |
| 2009 | 8,403.00 | 0.52 | 0.50 | -0.08 | 0.18 | 0.51 | 0.85 | 1.15 |
| 2008 | 8,190.00 | 0.49 | 0.52 | -0.11 | 0.15 | 0.48 | 0.84 | 1.14 |
| 2007 | 8,621.00 | 0.44 | 0.58 | -0.21 | 0.11 | 0.47 | 0.82 | 1.12 |

**Table 14:** Freshmen PSU (Raw) Scores and Enrollment in PrivateTeacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 5,573 | 509 | 68 | 414 | 472 | 516 | 552 | 590 |
| 2017 | 4,928 | 510 | 70 | 410 | 472 | 516 | 554 | 592 |
| 2016 | 7,620 | 487 | 72 | 393 | 444 | 490 | 533 | 576 |
| 2015 | 7,093 | 490 | 75 | 388 | 446 | 494 | 539 | 581 |
| 2014 | 7,148 | 496 | 74 | 396 | 452 | 499 | 542 | 587 |
| 2013 | 8,255 | 493 | 73 | 395 | 450 | 496 | 540 | 584 |
| 2012 | 11,190 | 494 | 74 | 397 | 450 | 496 | 541 | 586 |
| 2011 | 11,760 | 497 | 74 | 396 | 452 | 500 | 543 | 594 |
| 2010 | 12,036 | 488 | 67 | 398 | 452 | 493 | 530 | 566 |
| 2009 | 10,823 | 484 | 67 | 398 | 444 | 488 | 527 | 565 |
| 2008 | 10,295 | 475 | 69 | 385 | 433 | 480 | 520 | 557 |
| 2007 | 9,449 | 474 | 70 | 380 | 429 | 481 | 522 | 558 |

**Table 15:** Freshmen PSU Z-Scores and Enrollment in Private Teacher Colleges 2007-2018

| | N | mean | sd | p10 | p25 | p50 | p75 | p90 |
|---|---|---|---|---|---|---|---|---|
| 2018 | 5,573.00 | 0.04 | 0.63 | -0.83 | -0.31 | 0.10 | 0.43 | 0.78 |
| 2017 | 4,928.00 | 0.08 | 0.65 | -0.84 | -0.27 | 0.15 | 0.49 | 0.83 |
| 2016 | 7,620.00 | -0.13 | 0.66 | -1.00 | -0.53 | -0.10 | 0.29 | 0.69 |
| 2015 | 7,093.00 | -0.12 | 0.68 | -1.06 | -0.52 | -0.08 | 0.33 | 0.71 |
| 2014 | 7,148.00 | -0.07 | 0.67 | -0.99 | -0.47 | -0.04 | 0.36 | 0.76 |
| 2013 | 8,255.00 | -0.10 | 0.66 | -0.99 | -0.48 | -0.07 | 0.33 | 0.74 |
| 2012 | 11,190.00 | -0.09 | 0.67 | -0.97 | -0.48 | -0.07 | 0.34 | 0.75 |
| 2011 | 11,760.00 | -0.04 | 0.68 | -0.95 | -0.44 | -0.01 | 0.38 | 0.85 |
| 2010 | 12,036.00 | -0.12 | 0.61 | -0.94 | -0.45 | -0.07 | 0.27 | 0.60 |
| 2009 | 10,823.00 | -0.15 | 0.61 | -0.93 | -0.52 | -0.11 | 0.24 | 0.59 |
| 2008 | 10,295.00 | -0.23 | 0.63 | -1.05 | -0.62 | -0.18 | 0.18 | 0.52 |
| 2007 | 9,449.00 | -0.24 | 0.63 | -1.10 | -0.65 | -0.18 | 0.19 | 0.52 |

**Table 16:** Institutions' Value Added: Balance

| | Full sample | | Below | | Above | | Diff. | |
|---|---|---|---|---|---|---|---|---|
| | Mean/SD | Obs. | Mean/SD | Obs. | Mean/SD | Obs. | Mean | t-stat |
| ED Score | -0.00 | 31,417 | 0.01 | 9,100 | -0.03 | 9,303 | 0.036* | 2.456 |
| | (1.00) | | (1.00) | | (0.99) | | | |
| Female | 0.71 | 31,417 | 0.72 | 9,100 | 0.72 | 9,303 | 0.001 | 0.209 |
| | (0.45) | | (0.45) | | (0.45) | | | |
| Private HS | 0.06 | 30,804 | 0.07 | 8,961 | 0.06 | 9,166 | 0.008* | 2.199 |
| | (0.25) | | (0.26) | | (0.24) | | | |
| High SES HS | 0.17 | 25,909 | 0.17 | 7,648 | 0.17 | 7,936 | 0.002 | 0.384 |
| | (0.37) | | (0.37) | | (0.37) | | | |
| Experience | 15.94 | 31,417 | 16.13 | 9,100 | 14.32 | 9,303 | 1.803*** | 16.131 |
| | (7.67) | | (7.41) | | (7.75) | | | |
| Distance to cutoff | 4.20 | 31,417 | -12.24 | 9,100 | 10.54 | 9,303 | -22.786*** | -216.946 |
| | (35.74) | | (7.09) | | (7.16) | | | |
| Math Test Score (PAA) | 2.66 | 31,417 | 2.70 | 9,100 | 2.55 | 9,303 | 0.150*** | 13.188 |
| | (0.80) | | (0.77) | | (0.78) | | | |
| Reading Test Score (PAA) | 2.67 | 31,417 | 2.68 | 9,100 | 2.61 | 9,303 | 0.068*** | 6.732 |
| | (0.70) | | (0.68) | | (0.69) | | | |
| PAA year | 1987.44 | 31,417 | 1987.23 | 9,100 | 1989.21 | 9,303 | -1.979*** | -16.961 |
| | (8.02) | | (7.67) | | (8.15) | | | |
| ED year | 2007.54 | 31,417 | 2007.56 | 9,100 | 2007.62 | 9,303 | -0.067** | -3.100 |
| | (1.47) | | (1.46) | | (1.46) | | | |
| Number of Applications | 3.09 | 31,417 | 4.05 | 9,100 | 2.17 | 9,303 | 1.882*** | 71.112 |
| | (2.09) | | (1.96) | | (1.60) | | | |
| Admitted | 0.51 | 31,417 | | | | | | |
| | (0.50) | | | | | | | |
| Observations | 31,417 | | 9,100 | | 9,303 | | 18,403 | |
| Teachers | 15,939 | | 5,657 | | 9,303 | | 11,086 | |

# Figures

**Figure 22:** Area Under the Receiving Operators Curve

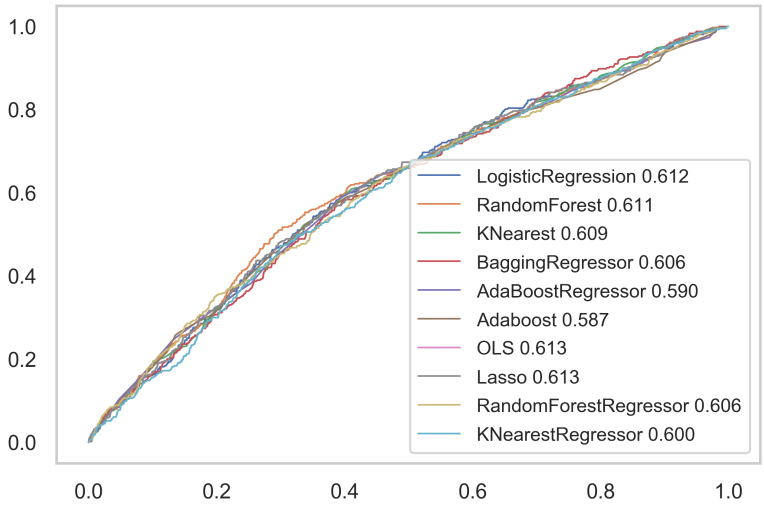
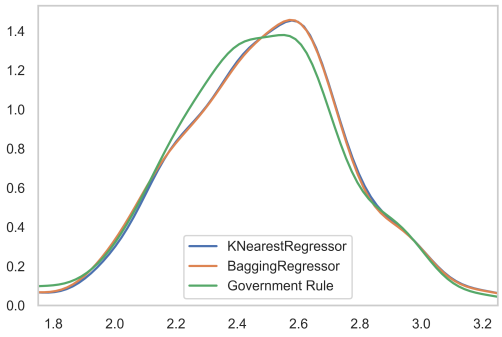
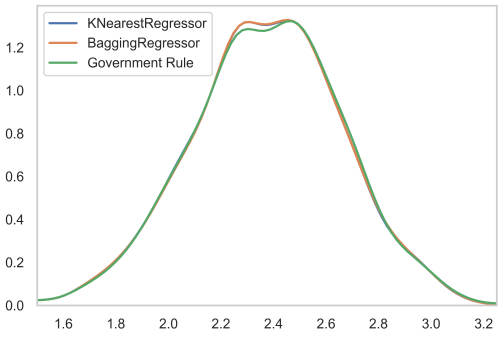



**Figure 23:** Distributional change in portfolio score for predicted good and bad performers

**(a)** Good Performers



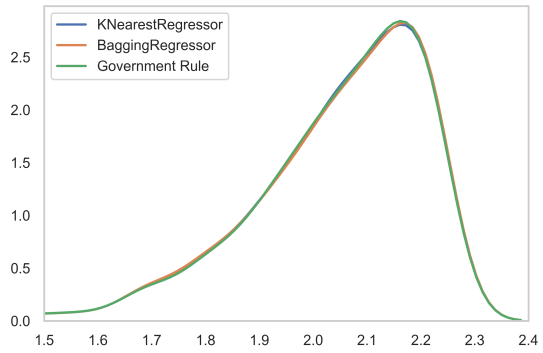

**(b)** Bad Performers



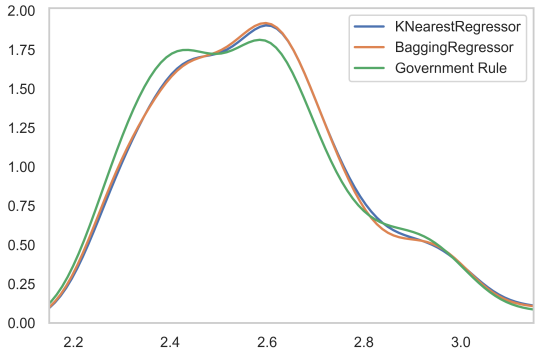

Note: .

**Figure 24:** Distributional change in portfolio score by prediction group

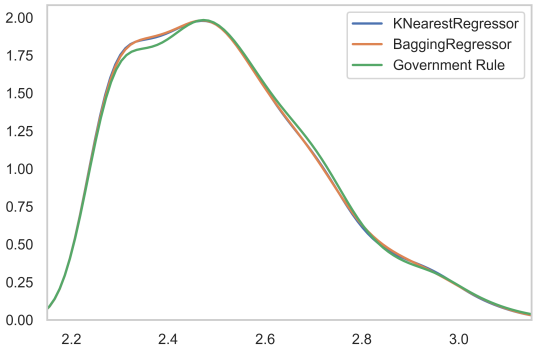
**(a)** True Positives



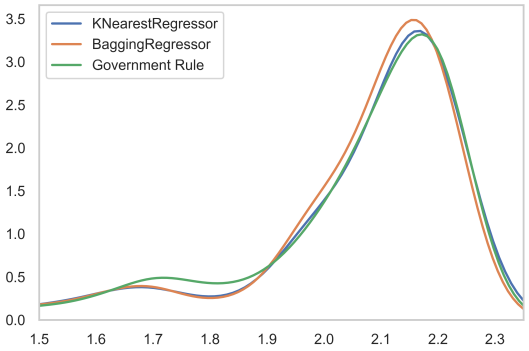

**(b)** True Negatives




**(c)** False Positive




**(d)** False Negatives




Note: Figures ...

**Figure 25:** Data contribution to the model

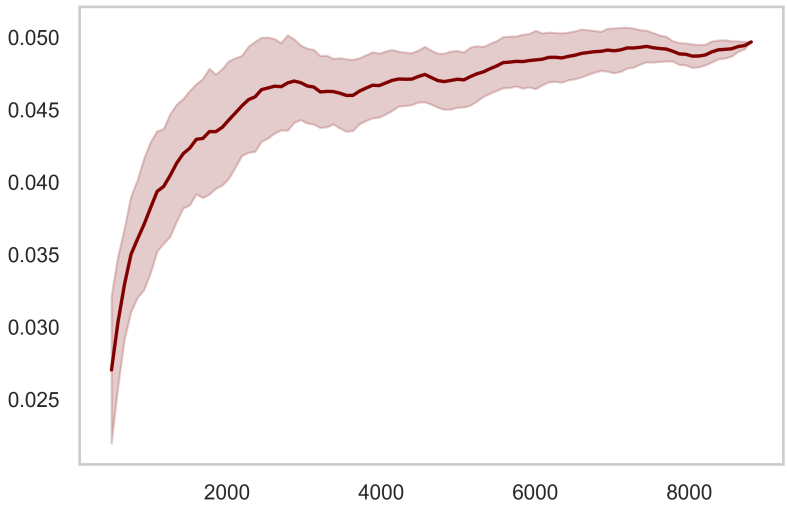
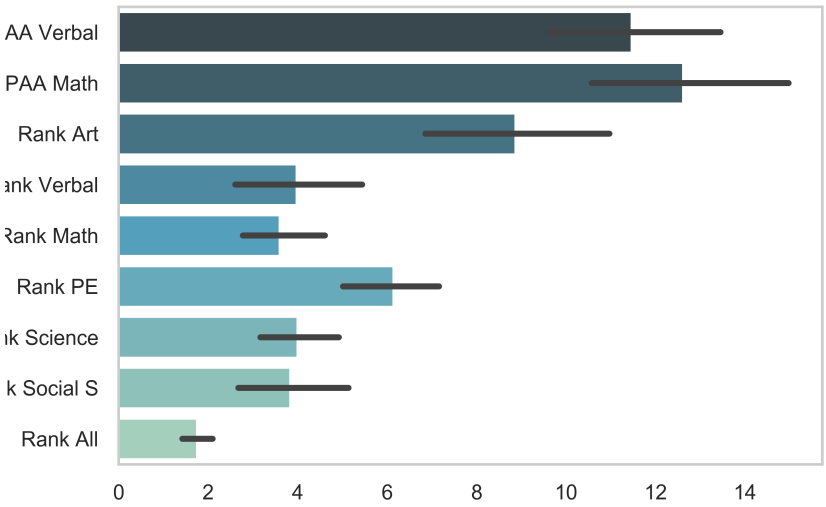



**Figure 26:** Variable contribution to the model



# Appendix

## Data Description

The information on the population of teacher comes from administrative records maintained by the Ministry of Education of the Government of Chile (MINEDUC).[10] The data, available for the period 2004-2017, contains 3.4 million records (an average of about 215K teachers per year), which correspond to 370,000 unique teachers. This dataset has information on demographics (gender, age), specialization level (preschool, primary, secondary) and subject (math, language, general, other), of the teacher, as well as tenure (years working as a teacher) and job characteristics (hours of contract, type of contract and starting in 2008, wages). The data on college entrance exams comes from a newly digitalized historical dataset (Gallegos and Neilson (2017)). The data amount to a total of 6.5 million records over the 1967 to 2016 period, which account for 99 percent of the published aggregate figures. About 70 percent of the 370K teachers are matched to their college entrance exam scores. We merge these data to each of our outcome measures:

**Table 17:** Teacher Productivity Measures

| Outcome Measure | Observations | Teacher Average Age | Year Collected |
|---|---|---|---|
| Graduation from Teacher Colleges | 84,847 | 23.1 | 2009-2017 |
| Exit Exams | 35,355 | 25.4 | 2009-2017 |
| Wages in Schools | 117,141 | 36.9 | 2011-2017 |
| Employment in Schools | 240,549 | 37.2 | 2003-2018 |
| Government Evaluations | 63,539 | 40.4 | 2004-2017 |
| Value Added and Students' Perceptions$^{(a)}$ | 3,756 | 44.9 | 2016-2017 |

Note: Table 17 lists our host of teacher productivity measures, with their number of observations, age at which the outcome was measured and year of data collection. $^{(a)}$ the number of observations for value added and students' perceptions correspond to the number of students

*Exit Exams*. The main objective of the teacher exit exam was to assess[11] the qualification of recent teacher graduates. The exam, first implemented in 2008, was gradually expanded by year and by the level at which teachers specialize (i.e. pre-school, primary and secondary). The information produced by the exit exam was thought to inform teacher colleges, policy makers and the test-takers themselves, but there were no associated consequences to its results. The teacher exit exam was called *INICIA* between 2008 and 2015, and was replaced by a new exit exam, called *Diagnostica* in year 2016. Like *INICIA*, the *Diagnostica*'s results do not have consequences. However, as opposed to *INICIA*, taking the *Diagnostica* exam is required to graduate from teacher colleges.

*Government Evaluations*. The Teachers' Evaluation (*Evaluación Docente* in Spanish, or ED onwards) assesses classroom teachers who work in public schools. The ED, implemented[12] gradually

[10] At the beginning of each academic year the MINEDUC requires all schools from the country to provide information on all their teachers.

[11] The exam measured (1) disciplinary knowledge (e.g. math knowledge for math teachers); (2) pedagogical knowledge (capacity of explaining concepts in a coherent way); (3) writing skills, and (4) capacity to use ICT (information and communication technologies) for teaching purposes.

[12] Our Online Appendix describes the ED implementation in detail. The assessment is composed by four components, with different weights: (i) a self-evaluation questionnaire (10%); (ii) a third-party reference report, filled by the school principal or supervisor (10%); (iii) one peer review (20%), and a *teacher performance portfolio* (60%). The portfolio component aims to collect direct evidence on teaching skills, pedagogical decisions and classroom practice. It includes two modules. In the first module, teachers plan a class defining its contents and related assessments. They are also asked questions about teaching practices. The second module consists in a videotaped class followed by a question-

since 2003, classifies teachers in four categories: unsatisfactory, basic, competent or outstanding. As opposed to the exit exam results, the ED results have consequences. Teachers classified in the 'competent' or 'outstanding' categories can opt to receive a monetary bonus.[13] Teachers classified in the unsatisfactory level are required to retake the ED. If after three attempts they remain in the unsatisfactory category, teachers must leave their schools and retire.

*Administrative Data*. We made significant efforts with MINEDUC to collect further administrative data to produce measures of teacher productivity at different moments of their careers. We gathered historic information of initial enrollment in teacher colleges, and data on graduation up to eight years after initial enrollment. In addition, since 2004 the MINEDUC keeps administrative records with information on the population of teachers working in schools.[14] With these sources of information, we build measures of graduation conditional on enrollment in teacher colleges, probability of working in schools conditional on graduation, and wages conditional on working in schools.

*Value Added*. Despite its rich administrative records, Chile does not implement tests designed to measure value added. To overcome this lack of data we implemented those tests for a sample of 9th, 10th and 11th graders and their teachers. The tests measured math skills at the beginning and the end of the academic year, and were specially designed to measure the gain in student achievement during that period.[15]

*Student Perceptions*. As part of the same project described above, we collected information from students' perceptions regarding effective teaching. The perceptions survey contains 38 questions and follows the recommendations of the Measures of Effective Teaching study carried out in the U.S. (Kane and Cantrell, 2010). Questions are categorized eight dimensions of teaching practices and classroom environment: positive culture and learning environment, student understanding checked for and ensured, engaging learning environment, expectations held by teacher, student input and ideas valued, learning fully internalized by students, encouraging and supporting relationships fostered, and classroom participation. We use students' perceptions in a as an additional measure of teacher quality.

*Teacher Wages in Voucher Schools*. Teachers from the voucher sector are not subject to any mandatory evaluation. However, while teachers working in public schools benefit from a special labor code, voucher teachers work under the regular, more flexible, labor code. Therefore their wages are more likely to be related to productivity, reflecting how the labor market values their performance.[16] The Ministry of Education keeps administrative records on the wages paid by schools to teachers, which we use as a proxy for the voucher teachers quality.

naire on the students behavior and understanding, and the teacher's own performance. The ED classifies teachers using a weighted score based on the components (i) to (iv) above.

[13]This system is complemented with a series of certification programs associated to relatively small monetary rewards, known as AVDI and AEP, which also started about a decade ago. AVDI stands for Asignacion Variable por Desempeno Individualin Spanish, while AEP stand for Acreditacion de la Asignacion a la Excelencia Pedagogica. More information on these programs can be found at: http://www.docentemas.cl, http://www.avdi.mineduc.cl y http://www.aep.mineduc.cl.

[14]At the beginning of each academic year the MINEDUC requires all schools from the country to provide information on all their teachers.

[15]The tests were developed by the Measurement Center of the Catholic University of Chile (MIDE UC) and are called SEPA tests. The name SEPA comes from the acronym in Spanish of Learning Progress Assessment System (*Sistema de Evaluacion de Progreso del Aprendizaje*. In De Gregorio et al. (n.d.) we use these learning gains to study the effectiveness of traditional teachers vis-a-vis "Ensena Chile" teachers. "Ensena Chile" teachers are recent college graduates recruited to teach for two years in low-income communities, resembling the U.S. experience of "Teach for America" teachers.

[16]This is generally not possible in most countries where teacher labor markets are heavily regulated. Chile adopted in 1981 an innovative nationwide school voucher system for primary and secondary education that still operates today. A description with focus on the teacher working regimes can be found in (Mizala and Schneider, 2014).

## Government Policy

The Chilean Government enacted in 2017 the Law 20,903, which creates a new system of professional development for teachers in the country. The Law includes guidelines for the recruitment, development and retention of teachers.

### Admission to Teacher Colleges

The Law included new conditions for teacher colleges' admissions, making use of the PSU score distribution and High School GPA ranking, defined as follows:

The PSU score for a given year is equivalent to the average score in Mathematics and Language, considering all test takers on a particular year. The Agency in charge of the PSU test, DEMRE will deliver the necessary statistics regarding the PSU scores and a certificates documenting that the score for each applicant.

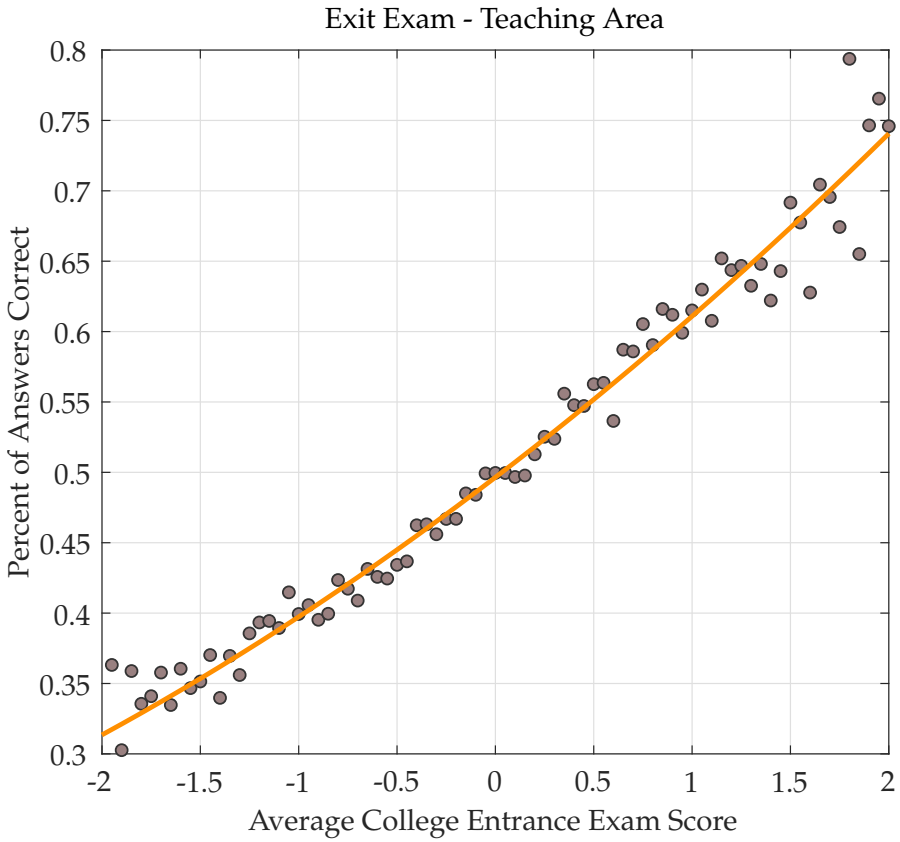
The High School GPA ranking is computed by the Ministry of Education, MINEDUC, taking into consideration the high school GPA of all students in the same cohort in the respective high school. Through the website Ayuda Mineduc (Mineduc Help), the Ministry will provide a document certifying the applicants' High School ranking.

The requirements for admissions in teacher colleges are designed to be gradually stricter over time:

- For the admission process in years 2017 to 2019, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 50th percentile of the distribution
  - Achieved High School GPA in the top 30% of performance.
- For the admission process in years 2020 to 2022, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 60th percentile of the distribution
  - Achieved High School GPA in the top 20% of performance.
  - Achieved a PSU score that is at least as high as the 50th percentile of the distribution **and** a High School GPA in the top 40% of performance.
- For the admission process in year 2023 and onwards, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 70th percentile of the distribution
  - Achieved High School GPA in the top 10% of performance.
  - Achieved a PSU score that is at least as high as the 50th percentile of the distribution **and** a High School GPA in the top 30% of performance.

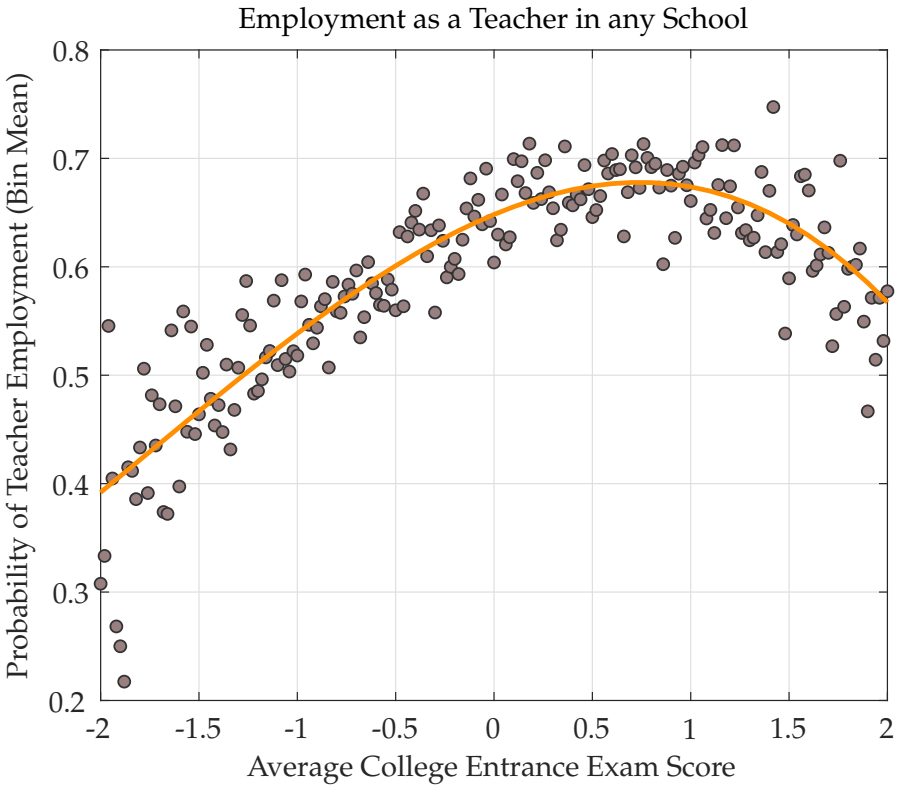
All of the conditions stated above are designed as **minimal** requirements for admission to teacher colleges. Each institution is allowed to consider stricter conditions, define number of vacancies or slots and application mechanisms. However, all the requirements must be informed before the beginning of admission process, each year.

**Figure 27:** Estimated Conditional Mean of Teaching College Exit Exams



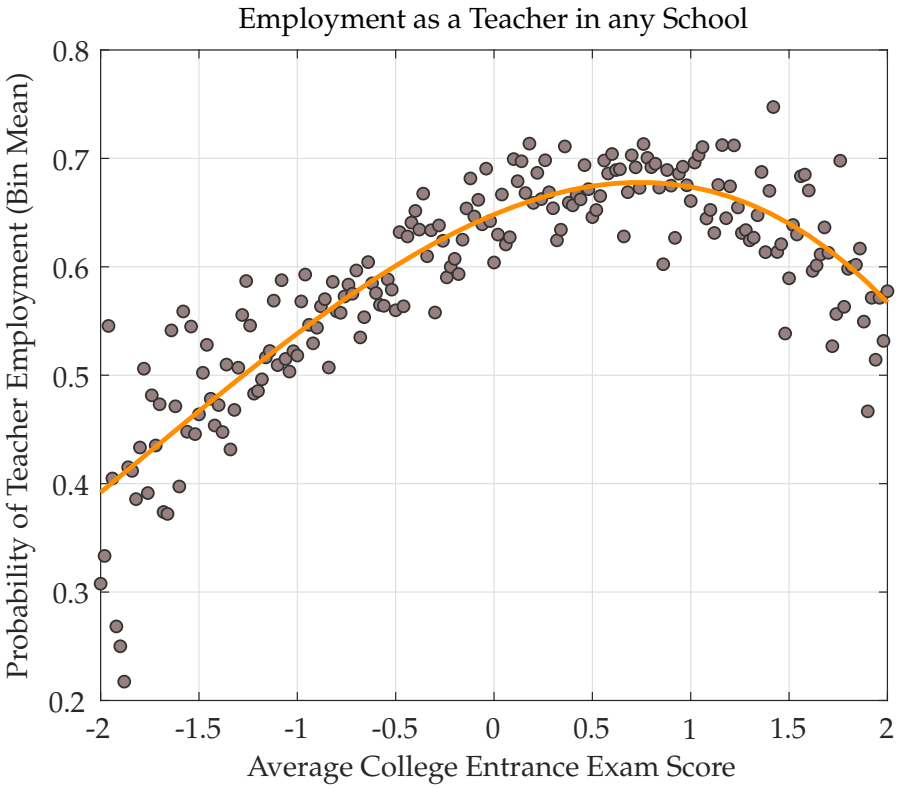

Note: This figure shows the binned plot of the mean of teacher college exit exams conditional on pre-college academic achivement measured by the average of standardized math and language scores. See Table 1 for a summary of the correlation of pre-college academic achievement and other measures of teacher productivity.

**Figure 28:** Estimated Probability of Being a Teacher at Any School




Note: This figure shows the binned plot of the estimated probability of a graduate of a teaching college being a teacher at any type of school five years after graduating. See **??** for a summary of the correlation of pre-college academic achivement and other measures of teacher productivity.

**Figure 29:** Estimated Probability of Being a Teacher at Any School




Note: This figure shows the binned plot of the estimated probability of a graduate of a teaching college being a teacher at any type of school five years after graduating. See **??** for a summary of the correlation of pre-college academic achivement and other measures of teacher productivity.

## Additional Information About *Beca Vocacion Profesor*

**Figure 30:** Marketing Material used by Ministry of Education

