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

December 31, 2019

Latest Version Downloadable: here

This paper studies screening and recruiting policies that use pre-college academic achievement to restrict or incentivize entry to teacher-colleges. Using historical records of college entrance exam scores since 1967 and linking them to administrative data on the population of teachers in Chile, the paper first documents a robust positive and concave relationship between pre-college academic achievement and several short and long run measures of teacher productivity. We then evaluate 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 at the individual level as well as the effect on the overall stock of teachers predicted effectiveness. We then evaluate the effects of a recent screening policy that forces teacher colleges to exclude below-average students. We quantify the policies effectiveness by retroactively simulating the policy rule and evaluate its success at screening out low performing teachers and mistakenly high 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. In both policies studied, screening low performing students is more effective than targeting recruitment efforts to only very high achieving students. Taken together, these findings suggest that the combination of better administrative data and flexible prediction methods can be used to implement practical screening and recruiting policies in some contexts and allow for better targeting of investments in future teachers.

*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 NBER Education Fall Meeting, 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) and dedicates this paper to the memory of Robert J. LaLonde. 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

Effective teachers matter for students' short and long run outcomes (Chetty et al., 2014). Accordingly, a common policy objective of governments all over the world is to increase the productivity of the stock of teachers in classrooms (OECD, 2005). These policies can be classified into those that look to increase the effectiveness of teachers once they are in the classrooms through incentives, training, accountability measures or rewards. An alternative set of policies that are less studied are aimed at recruiting or screening candidates before they enter teacher colleges or the teaching profession (Jackson et al., 2014).

Much of the past research has focused on the first set of policies. Recruiting and screening policies can be convenient compared to on-the-job policies for several reasons. The first is these policies can prevent students from exposure to ineffective teachers since it is usually difficult to remove low-performing teachers once employed. Second, it is logistically and politicaly hard to implement pay for performance schemes that look to incentivize effort (Hoxby, 1996; Hanushek, 2011). Third, there is little evidenece that later investments in training have a significant influence on teacher productivity (Jackson, 2012; Lombardi, 2019). Finally, targeting investments on the set of students who have the highest chance of being effective as a teacher later on can allow for more efficient use of resources. However, successful recruiting and screening policies are only possible to design if predicting teachers effectiveness ex-ante is feasible, something that has been elusive in the past (Harris and Sass, 2011; Jackson et al., 2014). Importantly, the design of effective recruiting policies depends crucially on the availability of data on the determinants of future teacher effectiveness. Governments have historically lacked this kind of information and there is scarse evidence that the data that does exist can reliably predict teacher effectiveness before entering higher education.

Data is now becoming more abundant as administrative sources become available; historical records are being digitalized and new information is getting recorded in more detail than ever before (Figlio et al., 2017). Combined with the development of improved algorithms, these data are lowering the cost of making increasingly accurate predictions and are influencing decisions, such as hiring, in many markets (Agrawal et al., 2018; Chalfin et al., 2016). These trends renew the interest and potential for screening and recruiting policies to be utilized by policy makers and have begun to be implemented in several countries.

In this paper we use we use recently digitalized historical records from 1967 onward to document the relationship between teachers' own academic achievement at age 18 and several measures of teacher productivity up to 30-40 years later. Second, we use the centralized college assignment mechanism to test causally whether access to more selective teacher colleges' explains these correlations finding null effects of college selectivity and teacher effectiveness. Having established a robust relationship between pre-college academic achivement and later teacher effectiveness, we then study two recent policy implementations that use pre-college achievement to recruit or screen out students entering teacher-colleges. We use a RDD produced by a recruitment policy to confirm that the predicted relationship between pre-college academic achievement and teacher productivity is invariant to recruiting policies in the short and medium run. We then use the policy changes together with our predictive model of teacher effectivenss to evaluate the impact of both recruiting and screening policies. Finally, we use standard machine learning methods together with data on entrance exams and rich high school transcript data, to show that a data driven screening policy can improve upon the performance of the current simple linear screening policy rule.

Our first set of findings show that there is a robust positive and concave relationship between teachers' pre-college academic achievement and a variety of short, medium and long run teacher outcome measures. Teachers' short and medium run outcomes include the probability of graduation from teacher colleges, college exit exams, and employment and wages in schools. Long run measures of productivity include government teacher evaluations, student test scores, and school value added as well as own teacher value added. Broadly, we find that below average pre-college achievement is systematically associated with lower performance as teachers measured up to thirty and forty years later.

Our second finding is that the observed correlation is caused by access to higher value added teacher colleges. We address this question directly by estimating teacher colleges' value-added using a regression discontinuity design building on institutional features of the Chilean centralized admissions system. Use data on the population of applicants to teaching colleges from 1977 to 2011, we find no evidence that any particular teaching college adds more value or contributes to closing or increasing the predicted gap in teacher effectiveness. This excersize verifies that college training is not enough to undo initial differences and that pre-college academic readiness has persistent relationship with later teacher productivity.

Given this evidence, we study two recent policies that used college entrance exams to screen out or recruit students entering teacher-colleges. The first policy, implemented in 2011, offered full tuition subsidies for high scoring applicants and also required participating institutions to reject low scoring students. We evaluate this *'carrots and sticks'* policy using a regression discontinuity based on the eligibility score cutoffs for high and low scoring applicants. Our findings show that the fraction of low scoring students decreased from 0.24 to 0.17 while the take-up of high-scoring students jumped from 0.13 to 0.17 at the respective cutoffs. Seven years later, early indicators of teacher productivity 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.[1] Impotantly, these medium run effects provide evidence that the predicted relationship betweeen pre-college academic achievement and medium run outcomes are policy-invariant in this context.

We then turn to study a second policy enacted in 2017 that used pre-college academic achievement as a direct screening policy. It prevented all teacher colleges to admit applicants with scores below the national mean We replicate the policy rule back in time to describe who would have been affected and whether the excluded students indeed become ineffective teachers later on. Partial equilibrium analysis shows that if implemented, these rules would have bound 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.

[1] In 2016 tuition was made free for many other programs due to a different, nation-wide policy. The same regression discontinuity shows that for the newer cohorts, the effectiveness of the policy was significantly diminished. These results are consistent with contemporaneous work by Castro-Zarzur et al. (2019) and Castro-Zarzur and Mendez (2019).

Finally, we compare the current government recruiting policy to a series of potential data-driven policy rules and find that even simple screening policies can identify a significant portion of ex-post low performing teachers. In particular, we train a standard model that classifies potential teachers based on a rich set of pre-college test information. Partial equilibrium analysis shows that our data driven rule would have increased the number of students graduating in time in around 5%, increased the number of teachers working after 7 years of being enrolled in college in 22%, and increased the number of teachers working in well performing schools in 20%.

These results are important because they have direct policy implications. If teacher effectiveness, or lack thereof, 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 become ineffective teachers.[2] This is particularly relevant because teacher labor markets are known to be inefficient (Neal, 2011; Gilligan et al., 2018), mis-allocation of talent can be widespread and in many cases (Bau and Das, 2018), and there is limited scope to sideline or retrain ineffective teachers once they are in the system, especially in the public sector (see, e.g., Estrada (2019) for the Mexican case and Bold et al. (2017); Svensson (2019) for seven African countries). Taken together, 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 and not on the extremely talented or students with extremly low levels of prior academic achivement.

We contribute to the literature on teacher quality and prediction. We see our results as consistent with the existing evidence on the topic from the US and developed countries (Rockoff, 2004; Rothstein, 2006). In the case of Chile, most of our ability to predict teacher effectiveness comes from very low achieving students who become teachers and this margin may not be relevant in more developed countries. This evidence is also consistent with recent cross country descriptive work by Hanushek et al. (2019), who find that in developed economies differences in teacher cognitive skills can explain significant portions of the international differences in student performance (measured by PISA scores). In addition, this analysis uses rich pre-college academic achievement for the population of teachers which may have not be available to researchers in the past. In this sense, our findings highlight avenues for further research in an increasingly data-rich environment where prediction is a key input to policy design (see, e.g., Mullainathan and Spiess (2017); Kleinberg et al. (2017)). Newer methods are being implemented to exploit increasing amounts of data, and we believe that empirical exercises similar to ours will be increasingly common in the near future. (Athey and Imbens, 2019; Athey, 2019; Sajjadiani et al., 2019).

[2] An important consideration are the equilibrium reaction of teacher labor markets to the changing composition of the supply of teachers. There is important work studying teacher sorting in the context of Chile by Tincani et al. (2016) and Tincani (2018) which models with survey data the sorting process. This paper complements this work by showing empirical evidence of the relationship between pre-college academic achievement and later outcomes.

2 Context, Policy and Data

2.1 Context

Chile is a middle income country that has reached low levels of teacher absenteeism and low student-teacher ratio, close to the levels displayed by OECD countries (World Bank, 2013). Teacher absenteeism is estimated at 5% (Paredes et al., 2015) which is much lower than other countries in earlier stages of development (e.g., Chaudhury et al. (2006) estimate an average of 19 % for Bangladesh, Ecuador, India, Indonesia, Peru and Uganda). The student-teacher ratio is about 20, which is the result of an increasing number of teachers and a stable population of students over time. The country 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[3] has increased from 125,000 in 2008 to 164,000 in 2018 (MINEDUC, 2019), which has led to a reduced student-teacher ratio (from about 26 to 19). With enough teachers in the classrooms, and high student enrollment rates (OECD, 2009), the policy focus switched in the last ten years to bring more qualified individuals to the teaching profession.

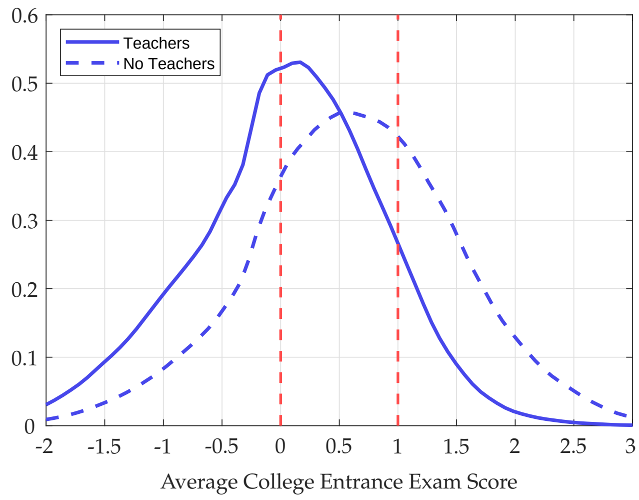
Attracting more skilled individuals to be teachers is challenging because, among other factors, teachers are typically paid less than comparable professionals (Mizala and Nopo, 2016).[4] Consistently, we know from the related literature that college graduates with higher college entrance scores are less likely to enter teaching (Manski, 1985; Hanushek and Pace, 1995; Vegas et al., 2001), and Chile is no exception. Figure 1 shows that between 2007 and 2010 students enrolled in fields other than education (engineering, law, medicine, etc.) scored about 0.6 standard deviation (σ) above the national mean in the college entrance exam,[5] while teacher college students scored only 0.1σ above. In addition, we computed that the scores from teacher colleges have been declining over time; in 1995 students from teacher colleges scored 0.3σ over the national mean. This pattern is similar to the evidence for the U.S. (Bacolod, 2006; Corcoran et al., 2004; Podgursky et al., 2004; Hoxby and Leigh, 2004).

[3]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.

[4]Mizala and Nopo (2016) estimate the earnings gap as the percentage of average earnings remaining after controlling for a set of characteristics linked to productivity. In Chile in particular, the underpayment for teachers was about 18% in 2007.

[5]These scores correspond to the average of the math and language exams. We describe the college entrance exam in section 2.3.

Figure 1: Distribution of College Exam Scores: Teachers Colleges vs Other Fields, 2007-2010



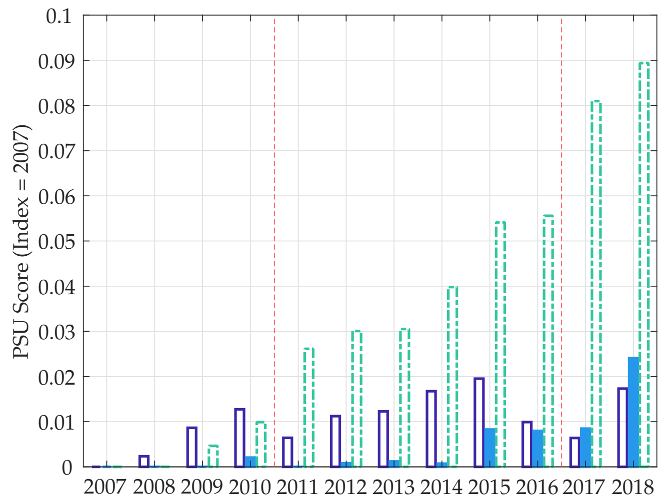
Note: Figure 1 plots the distribution of college entrance exam scores for freshmen in teacher colleges (continuous line) and freshmen in the health and STEM fields (dotted line), using data from 2007 to 2010. The entrance exam score (in standard deviation units) is the average of the math and language exams. We provide further details on the college entrance exam in section 2.3.

2.2 Recent Policies to Recruit Teachers In Chile

The Chilean government implemented two policies to recruit teachers in the last decade. The first policy, implemented in 2011, was the *Beca Vocacion Profesor (BVP)* and consisted in full tuition subsidies for prospective students who scored about 1σ above the mean in the college entrance exam. Importantly, the BVP policy also required participating teacher colleges to reject applicants with scores below the national mean. The second policy, that started in 2017, was a screening policy that imposed new requirements for admissions at all teacher colleges across the board. This government policy required applicants to teacher colleges to have college entrance exam scores at least as high as the median of the distribution of test-takers or have a high-school GPA in the top 30% of their high school graduating cohort.

Descriptive statistics suggest that both policies affected freshmen entrance scores, enrollment and the availability of programs in education versus other fields. 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 the percentage increase in PSU scores for freshmen in the education, health and STEM fields, from 2007 to 2018, taking 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 shows almost no variation in the scores achieved by freshmen in STEM or health fields.

Figure 2: Relative Freshmen College Exam Scores over Time, by Field

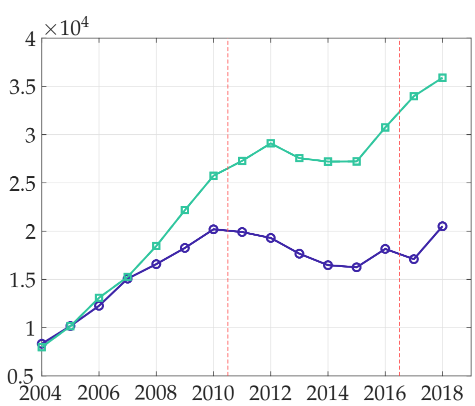


Note: In Figure 2 the dotted bar plots the percent change on the average college entrance exam score (labeled 'PSU Score') for freshmen in teacher colleges, the filled bar does the same for health, and the empty bar for STEM careers, from 2007 to 2018, using 2007 as the base year. The dotted vertical lines illustrate when the policy changes were introduced (BVP in 2011 and screening policy in 2017, respectively).

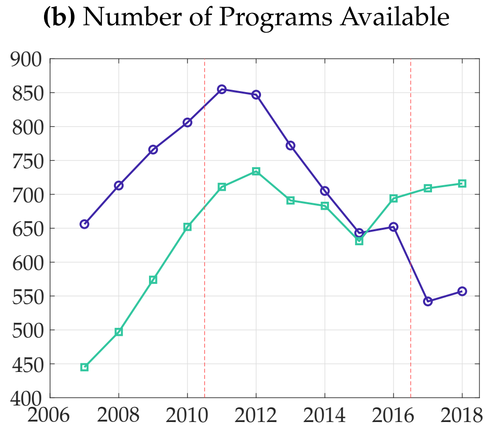
Figure 3 shows the evolution of enrollment (Figure 3a) and programs available (Figure 3b), for teacher colleges and health careers. Figure 3a suggests that the policy changes flattened the increasing trend of students enrolled in teacher colleges, while the trend continued for students enrolled in health programs. Consistently, Figure 3b shows that after 2011 the number of teacher college programs sharply declined, with a last and very steep reduction in 2017. In the case of health, for instance, the number of programs available maintained its increasing pattern with some plateau period between 2013 and 2015 but retaking the increasing trend towards 2018.

Figure 3: Enrollment and Programs Available over Time, by Field

(a) Number of Students Enrolled



(b) Number of Programs Available



Note: Figure 3 plots the evolution of enrollment (Figure 3a) and programs available (Figure 3b), for teacher colleges and health careers. The circles represent values for teacher colleges, while the squares do the same for health programs. The dotted vertical lines illustrate when the policy changes were introduced (BVP in 2011 and screening policy in 2017, respectively).

2.3 Data on Pre-College Academic Achievement

The main measure of teachers' pre-college academic achievement that we use in this paper is their scores on college 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.

The Chilean national college entrance exam is similar to the SAT in the Unites States. Currently, the exam is called the *Prueba de Selection Universitaria* (PSU) and has been administered once a year since 2004. Prior to that a similar test called *Prueba de Aptitud Academica* had been implemented from 2003 back to 1967, which makes Chile to have one of the longest running centralized college assignment systems in the world.[6] 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.

2.4 Data on Teacher Productivity

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, and external classroom teaching evaluations, all gathered from administrative records.[7]

In spite of the general richness in the data, what is not readily 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. (2019), where students took tests at the beginning and at the end of the academic year to capture value added of the teacher by controlling for initial conditions. As part of the same project we surveyed students on their perceptions regarding whether their teacher is effective.

We use all these sources of information to proxy teacher productivity merged with the digitized pre-college achievement described above. Below, we document each dataset. 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 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.

[6] 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).

[7] According to a recent review by World Bank, Chile has the most advanced system of teacher performance evaluation in Latin America (Bruns and Luque, 2015). The most important assessments are exit exams for graduates from teacher colleges and classroom evaluations.

Government Evaluations. We use information for 63K classroom teachers in public schools, evaluated between 2004 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, which makes wages grow with tenure and not expected to change with productivity. However, the voucher sector operates under the regular and more flexible labor code, and thus teacher wages can be given a market clearing interpretation, associated to productivity. They represent 62 percent of our sample.

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.[8]

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 most of the teacher productivity measures described in the previous

[8]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. (2019) 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.

section. We estimate parametric regressions of teacher outcomes at different moments of their careers on their own entrance exam scores taken at age 18. We also describe the empirical relationship showing non-parametric plots leveraging on our large sample sizes.

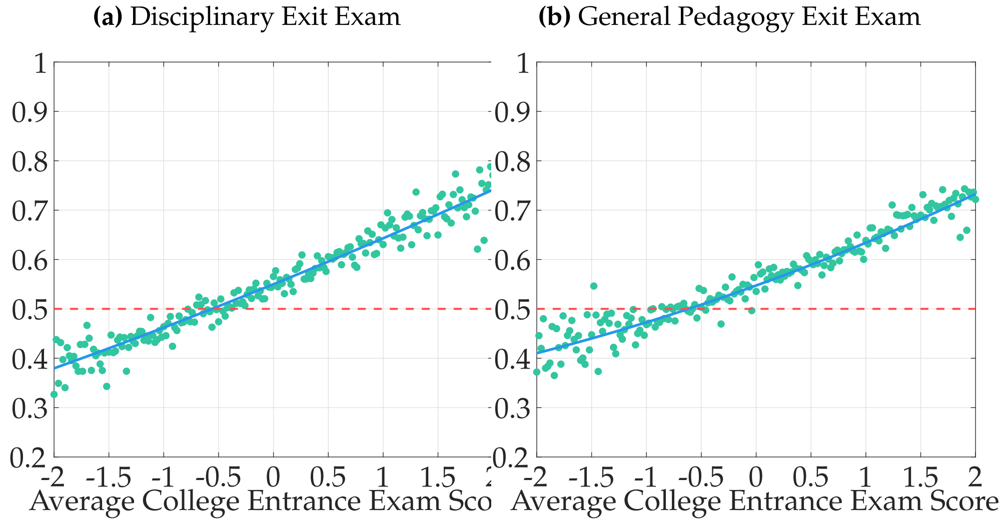
The general takeaway is that the empirical relationship between pre-college skills as a student and teacher productivity later on is positive and concave. In Table 1 we show the coefficients of 14 separate regressions for different measures of teacher performance on the college entrance exam score (in standard deviation units) and its square. The coefficients on scores are positive and significant, and most coefficients on the square are negative.

In Figure 4 and Figure 5 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. According to Table 1, one standard deviation on the test scores that current teachers took years ago, is associated to an increase of 0.50 σ on both the disciplinary and pedagogical skills in the exit exams respectively; and an increase in 0.46σ and 1.27σ in writing skills and ICT skills (available for smaller samples). The same pattern can be visualized in Figure 4 for the pedagogical and disciplinary tests. Non-parametric plots for ICT and writing tests are shown in the online appendix.

With regards graduation, the relationship appears concave for both graduation after 5 and 8 years after enrollment. The results show than an increase in one standard deviation on the college entrance exam scores leads to an increase in graduation rates after 5 (8) years of enrollment in 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 with potential higher expected income, or they might just have an attractive outside option in the labor market.

Figure 4: College Entrance Exam and Teacher College Exit Exams

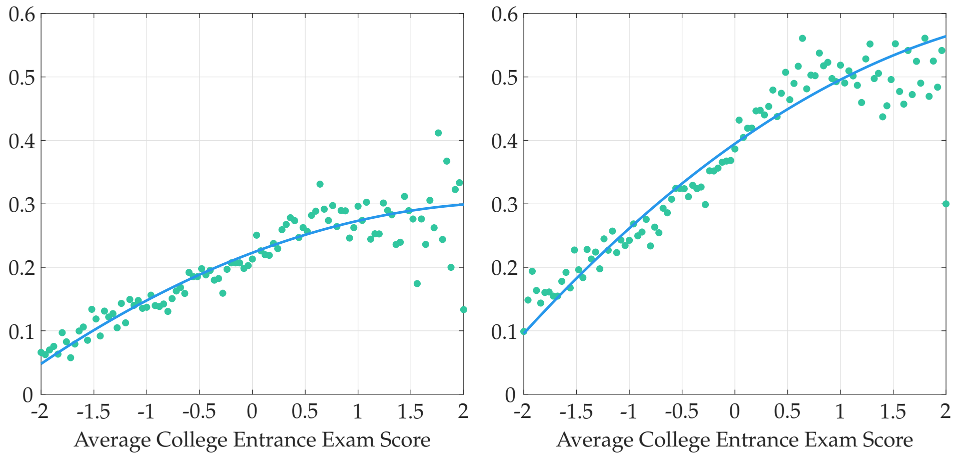
(a) Disciplinary Exit Exam **(b)** General Pedagogy Exit Exam



Note: The figures plot the fraction of correct answers in two subjects of the exit exam (Disciplinary in Figure 4a and Pedagogical in Figure 4b), within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in graduates who took the respective exit exam test between years 2009 and 2017. The sample sizes are $N = 35,355$ in Figure 4a, and $N = 33,409$ in Figure 4b.

Figure 5: College Entrance Exam and Graduation from Teacher Colleges

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

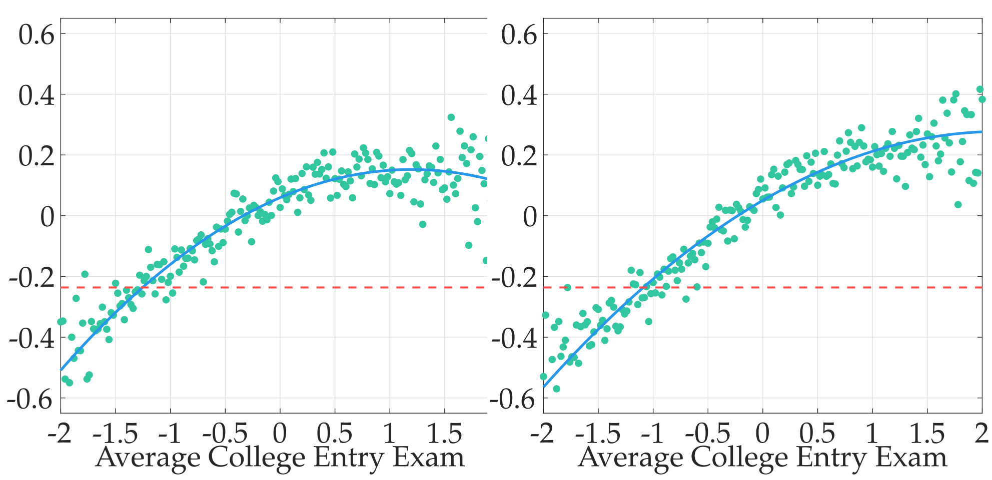


Note: The figures plot the probability of graduation after 5 years (Figure 5a) and 8 years (Figure 5b) of first enrollment, within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in students enrolled in years 2004 to 2009 who graduated between 2009 and 2017. In both Figures the sample size is of $N = 84,847$.

The next set of results are for later outcomes, when individuals are teaching and working in schools. Figure 6 show the bivariate relation between college entry exams scores and teacher evaluations taken up to 30 years later. The relationship is concave again, suggesting that early scores may have a higher potential for identifying low performance teachers than high performing ones thirty years later. Coefficients in Table 1 show that an increase of one standard deviation in entry exam scores translate into an increase of 0.62σ and 0.48σ on the teacher evaluation score overall and portfolio score respectively.

Figure 6: College Entrance Exam and In-Class Teacher Evaluation

(a) Overall Standardized Score **(b)** Portfolio Standardized Score



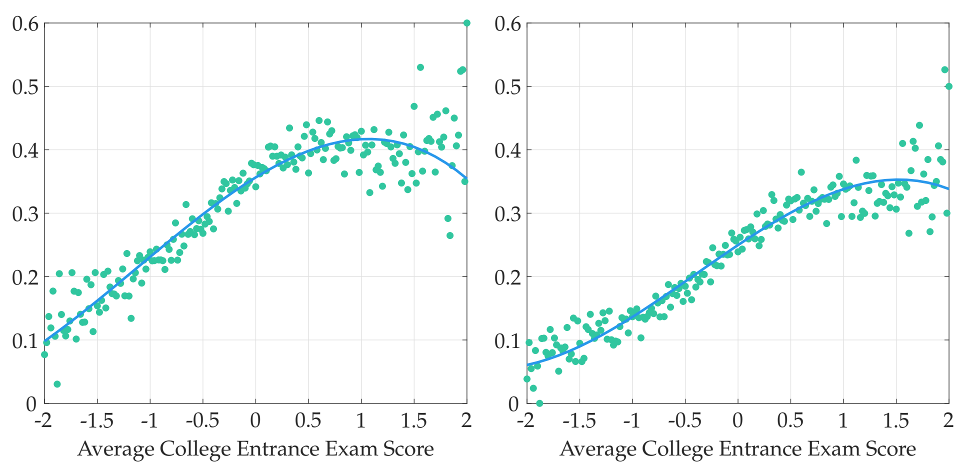
Note: The figures plot the teacher evaluation scores (overall in Figure 6a and the portfolio component in Figure 6b), within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in teachers evaluated between years 2004 and 2017. In both Figures the sample size is of $N = 63,539$.

Table 1 is consistent with the concave productivity story we were presenting in the Figures, and corroborates 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.

Figure 8 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. The change in wages in the private sector seem to be driven by both experience and scores, meanwhile for the public sector experience is the most relevant factor since salary increases ocurr in the base of seniority. Consistently, the coefficients in Table 1 show that a standard deviation increase in scores is associated to 0.26σ and 0.45σ 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.

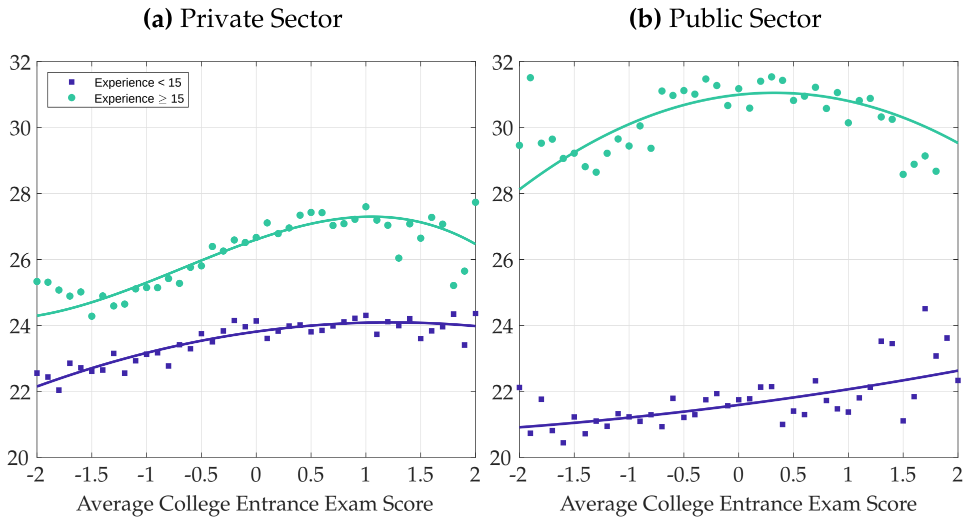
Figure 7: College Entrance Exam and Working in Schools

(a) Employment in Schools **(b)** Employment in Good Schools



Note: Figures 7a and 7b plot the fraction of teachers employed in schools within 100 equal-sized bins of the average college entrance exam score, and fits estimated lines using all the underlying data. The data consists in graduates from teacher colleges in years 1995 to 2017, who are employed (or not) between 2003 to 2018. In both Figures the sample size is $N = 240,549$.

Figure 8: College Entrance Exam Average and Wages (USD 2019)



Note: The figures plot the wages for teachers in public (Figure 8b) and private (Figure 8a) schools in US dollars (2019), within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using the underlying data. The data consists in wages reported by schools in year 2011. The sample sizes are of $N = 36,771$ in Figure 8b and $N = 58,523$ in Figure 8a.

Table 1: College Entrance Exam and Teacher Outcomes

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 in Schools	**Years after graduation 5 Years**	**10 Years**	**20 Years**	**Value 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: Table 1 reports results from 14 separate regressions of teacher outcomes on college entrance exam scores (labeled 'PSU Score') and its square. The PSU score is expressed in terms of standard deviations in all cases. The table is organized in four panels: graduation, exit exams, productivity measures and employment. All estimations include year and teacher specialization fixed effects Robust standard errors (in parentheses) are clustered by day of birth. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively.

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 a mandatory screening policy implemented in 2017 and requires accredited teachers' colleges to accept and allow enrollment only to students with entrance exam scores above the mean. 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.

4 A Carrot and Sticks Approach to Recruiting and Screening

This section presents results of the *Beca Vocacion Profesor* (BVP) policy, which we briefly described in subsection 2.2. 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. We provide details on specifics of the policy next, and then show results on college participation choices and students outcomes.

4.1 BVP Policy Specifics

The BVP was first implemented in the academic year 2011,[9] This policy gave full scholarships and other incentives such as stipends and paid semesters abroad for students with scores from approximately the highest 30% of the admissions test distribution, who matriculate at teaching colleges. 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,[10]. 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.

In addition, the policy imposed colleges to screen out low performing students. In particular, teacher colleges were required to implement a cutoff score at the 50th percentile of the average score distribution if they wanted their top students to matriculate with the BVP scholarship.[11]. 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 Descriptive Effects

On Aggregated] 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

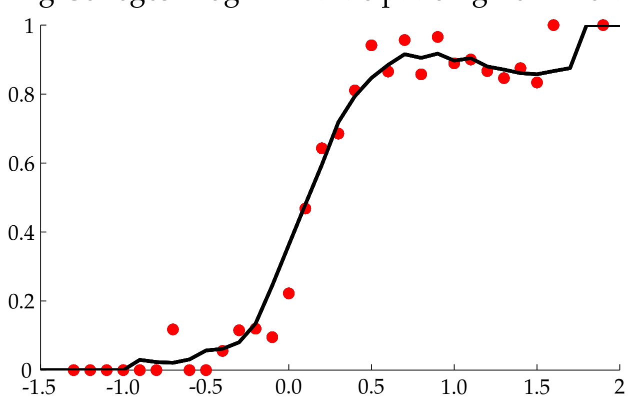
[9] Requirements for students included having applied and been admitted to an eligible teaching college as a new first year student in 2011 with an entrance exam score from December of 2010. Students previously enrolled in teaching careers are not eligible for this particular modality of the scholarship.

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

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

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 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 impact than it could have been due to low participation of teachers colleges, where only 50% decided to join the program. Figure 9 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 mandated 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 9: Teaching Colleges Program Participation given Average Score in 2010



Notes: Figure 9 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.

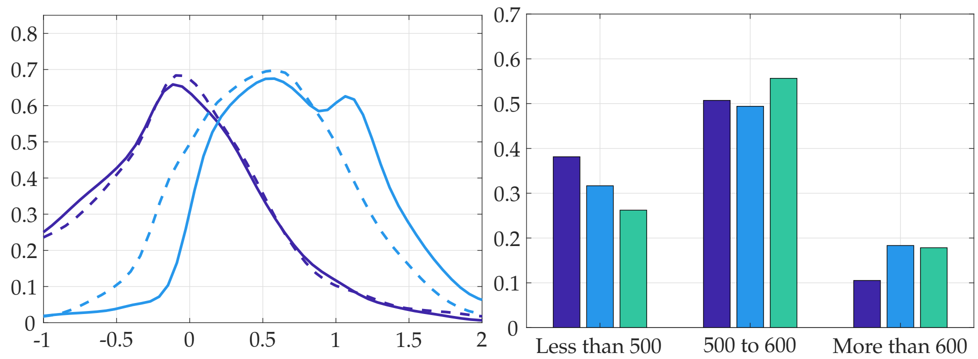
On Aggregated Students Choices. In the application process to the 2011 BVP, the number of applicants to the teacher scholarship was 28,179. Among these, only 3,385 (12%) were eligible for the benefit and 3,252 accepted it (96% of the elegibles).

Among the ones that accepted the fellowship, 3,063 are beneficiaries of the type I fellowship[12], which is the focus of this paper. The effect the policy had on the distribution of scores is quite striking. Figure 10 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 10 shows how the proportion of incoming teacher college students changed at specific intervals of interest.

[12]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 10: Aggregate Effects of *Beca Vocacion Profesor*

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



Note: In Figure 10a the continuous and dotted lines show the scores distribution for year 2010 and 2011 respectively. The ■ represents scores distribution for non BVP schools while the ■ shows the distribution for BVP schools.In Figure 10b ■ represent the shares in 2010 ■ represents shares for 2011 and ■ for 2017. Source: MINEDUC and DEMRE.

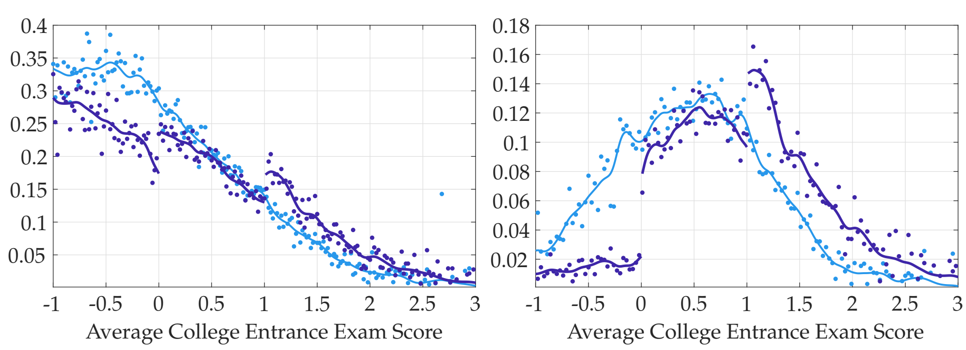
4.3 Causal Effects on Individual Students Choices

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.

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 11: $Pr(\text{Enrollment}|\text{Exam})$ at Teaching Colleges

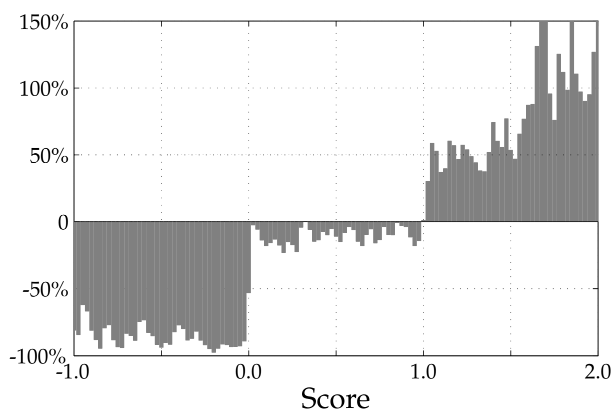
(a) $Pr(\text{Enrollment}|\text{Exam})$ in Any Teaching College **(b)** $Pr(\text{Enrollment}|\text{Exam})$ in Eligble Teaching Colleges



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. Color ■ represent the probabilities for 2011, while ■ the probabilities for 2010. Source: MINEDUC and DEMRE.

If we analyze only the probability of choosing an eligible teaching career, the probability for 2011 decreases under the average μ (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 12 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 μ (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 11. 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 12: Change in $Pr(\text{Enrollment}|\text{Exam})$ in Eligible 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.

4.4 Regression Discontinuity Design

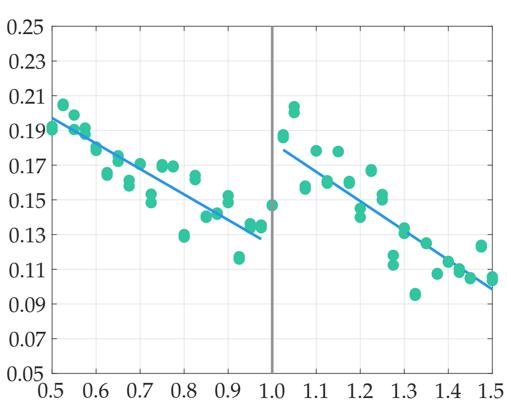
The program in their eligibility criterion uses a discontinuity that can be used to evaluate potential causal effects over different outcomes. The eligibility criterion is a decision rule that fix a cutoff point to select beneficiaries and non-beneficiaries for the program. This analysis allows estimating 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 outcome is a continuous non-parametric function. We will study the effect of the fellowship over the extensive margin by studying if the program atracted and therefore increased the number of teachers in the labor market with scores above the threshold of 600 points, and also if the financial incentives operated through the intensive margin by increasing the likelihood –conditional on scores– of graduating and entering early into the labor force as teachers for students with scores above 600, 700 and 720 points. This methodology does not use many assumptions, but is valid only locally around the threshold points.

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.

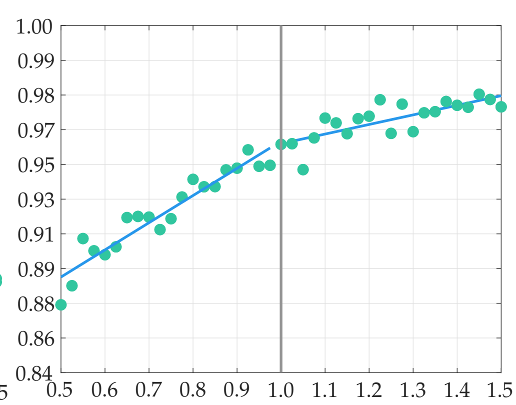
Graphically we observe the effect around the 600 points cutoff in Figure 13. The points are the average among students in groups of 5 points of difference and the line to each side of 600 is a linear regression. Figure 13b shows that the overall enrollment into universities didn't change for students with scores above the threshold; meanwhile Figure 13a shows that the threshold crossing effect increased the likelihood of enrolling into a teaching program from 12% to 18%. The combination of those results suggest that there was a shift in the margin of choosing a career rather than in the number of students that were qualified to apply for the scholarship in the margin. Students that would have prefered a non teaching career in the margin, ended enrolling into teaching colleges because of the financial incentive.

Figure 13: Effects of BVP over Enrollment

(a) Enrollment at Teacher Colleges



(b) Enrollment Overall



Note: Figure 13a plots the threshold crossing effect over the number of students enrollment into any teaching program conditional on PSU score while, Figure 13b shows the probability of enrollment in any program conditional on PSU score.

Although, we could not find any significant effect of the third level of benefits over the probability of enrollment in a teaching career we found important effects over the less generous incentives. Particularly, as Table 2 shows, the increase in the probability in 600 is of 5.3 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.

Table 2: RDD Regression across different BVP Thresholds

	Threshold = 600		Threshold = 700		Threshold = 720	
	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
Above Threshold	0.0538***	4	0.0262***	2.42	-0.007	-0.72
Constant	0.1204***	20.01	0.0283***	4.6	0.0033***	5.38
PSU	-0.0007***	-1.54	-0.0015***	-3.18	-0.005***	-1.12
N. Observations	18007		5450		4150	

4.4.1 *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 3: 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

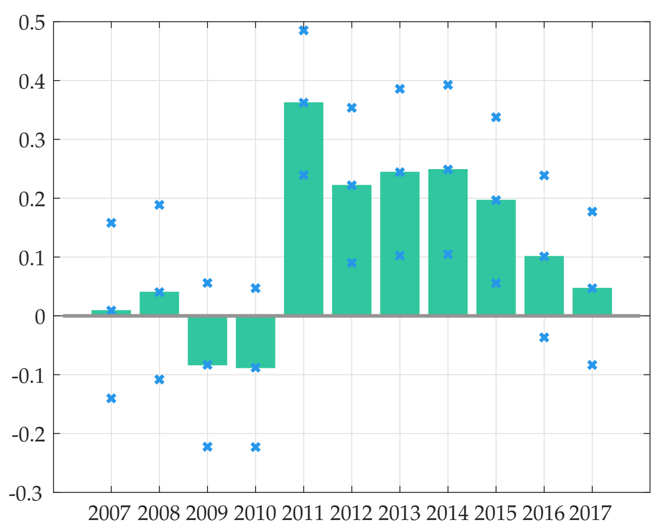
Note: Table 3 shows the results of estimating the equation $Y = \alpha_0 + \alpha_1 BVP + f(Score) + \mu$ where Y are the outcomes in each column, BVP is a dummy for being eligible for the BVB (i.e., having a college entrance exam score of 600 or above, and $f(score)$ is a function of the college entrance 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).

Moreover, as Figure 14 shows, there is an important contribution of the BVP policy on the extensive margin by marginally increasing the number of students with scores above the threshold in 20 students on average or 34%, in 24 the number of students working in schools after 7 years of enrollment or 52%, and in 20 the number of students working in good schools for those or 74%. However, it seems that the finantial incentive did not have an actual impact over the likelihood of succeeding in any of the outcomes mentioned before, conditional scores, as can be seen in all of the figures in the right panel. This means that even when the policy helped recruit more qualified students in the margin, the recruitment program turned being policy invariant because it did not have an effect on accelerating or improving the labor outcomes for those who received the fellowship.

Policy effects depend crucially on other policy

The results mentioned before depend however on the existence of other finantial restrictions outside teaching programs. 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 over enrolling into teaching programs. When examining the policy discontinuity over time, the threshold crossing effect over recruitment drops consistently until 2016 when gratuidad was implemented. Figure 15 plots the percentage increase in the freshmen enrollment rate in teaching colleges attributable to the threshold crossing effect until 2017. The exercise shows that the effect decreased considerably over time from around 35% in 2011 to almost zero percent in 2017.

Figure 15: RDD Estimates on Freshmen Enrollment in Teacher Colleges

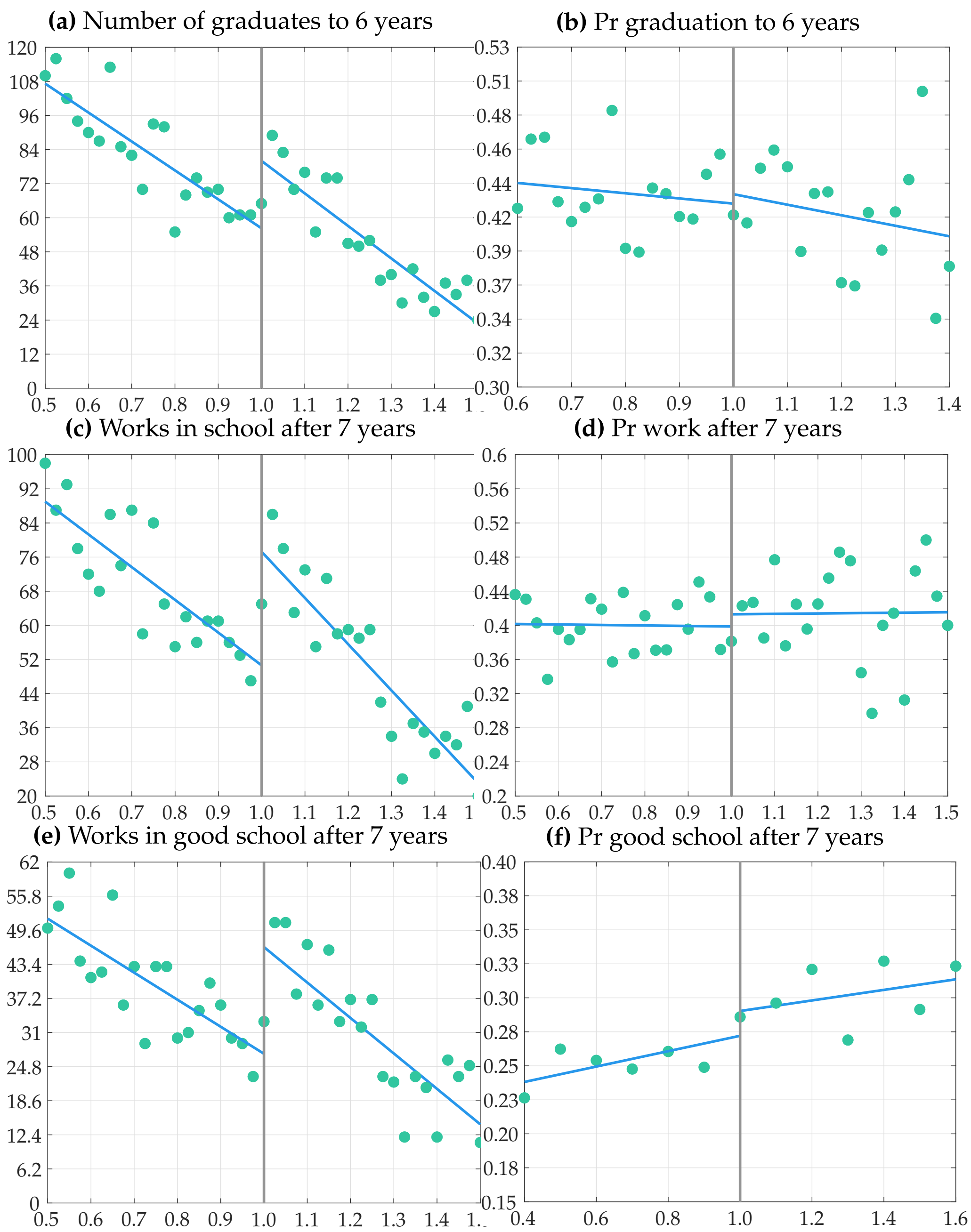


Notes: Figure 15 plots the coefficient α_1 over the average for each year (with 95% confidence intervals) estimated from the equation $Y = \alpha_0 + \alpha_1 BVP + f(Score) + \mu$ where Y is freshmen enrollment in teacher colleges each year, BVP is a dummy for being eligible for the BVP policy (i.e., 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 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. We estimate that 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 and screening restrictions are estimated to have reduced the bottom tail of the distribution by 4000. In terms of predicted productivity, the concavity of relation between pre-college academic achievement and predicted teacher effectiveness suggests that reducing low performers may be more effective than recruiting very high performers because of the uncertainty on the right tail of the distribution.

Figure 14: Effects of BVP over Enrollment



Notes: the figures graph a set of outcomes over PSU scores for all freshmen enrolled in 2011 who took the the college entrance exam in 2011 and where eligible for the BVP scholarship. In particular, figure Figure 14a plots the probability of enrollment in teacher colleges in 2011, Figure Figure 14b plots the probability of college graduation between March, 2013 and February, 2018, Figure Figure 14f plots the probability of employment in schools between 2013 and 2018; figure Figure 14c use data from years 2016 and 2017 to plot the pedagogical test scores conditional on taking the exam.

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)[13]. 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 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.

The second is to train and search among a family of flexible models a data driven method to screen out potentially low performing teachers by relying only on pre-college human capital characteristics. We evaluate how succesful our data driven method can be –compared to the screening criteria proposed by the government– at minimizing the mistakes of rejecting future effective teachers while mantaining fixed the number of ruled out students. Our hipothesis is that our data driven method may be able to explode non linearities existent in the data that could be useful for identifying bad performers that would have been ignored by a less flexible rule.

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 (P17) requires students to either have achieved 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 (P20), the 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 (P23), 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

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

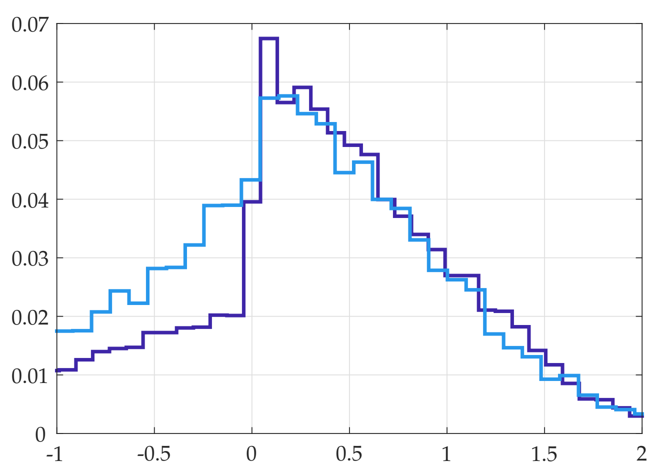
student is in the top 30% of the GPA distribution at their high school **and** has scores above the average, then that student can also matriculate at teaching colleges.

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

In 2017 the screening policy contributed to a significant reduction in the number of students admited whose entry exam score were below the cutoff of 500 relative to the population of students admited in 2016. As shown in Figure 16, the proportion of students with scores below the average reduced to half. However, some of the low scoring students were still elegible since their GPA ranking was compliant with government requirements. The government applied this rule to avoid potential unsatisfactory teachers to be admited into teacher colleges.

Figure 16: Score distribution for students enrolled in pedagogy



Note: In Figure 16, color ■ shows the score distribution for students enrolled in 2016, while ■ does the corresponding with students enrolled in 2017. The y axis are percentage of students in each bin from the total number of students in each the sample. Source: MINEDUC and DEMRE.

In order to evaluate the performance of the policy at ruling out bad teachers, we simulated the policy rule of 2017 backwards from 2007 to 2016. In Figure 17a we compare the number of students that would have been accepted against the share of students that would have been rejected if the policy was implemented, and results suggest that if the policy was implemented prior to 2017, the number of accepted students would have continue decreasing as well as the proportion of students that would have been rejected by the screening rule. According to the figure, the share of students with low ranking and low PSU were between 25 and 31% in 2007 and remained at this range until 2010. From 2011 onwards the trend decreased significantly due to the introduction of BVP and the countinuous recruitment of teaching schools into the BVP scheeme.

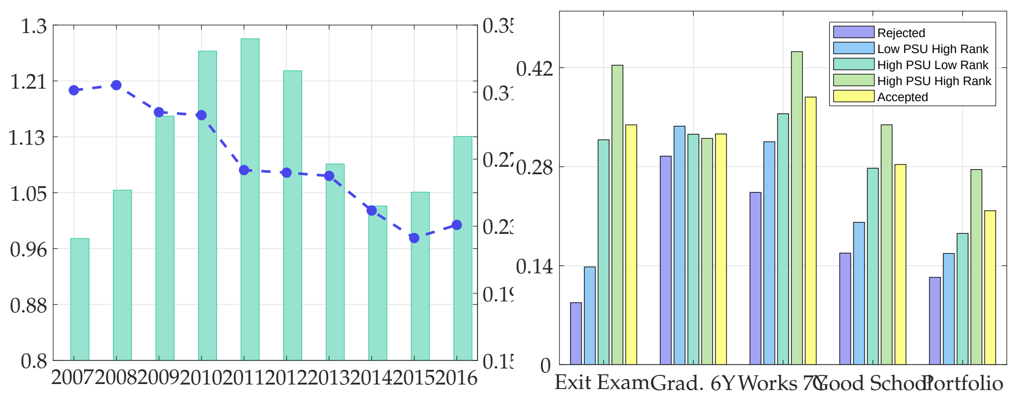
In Figure 17b, we compared labor outcomes of past teaching students classified for students that would have been rejected or accepted by the 2017 policy but from 2011 to 2016. According to our results in Figure 17b, students that would have been rejected by the 2017 screening policy performed remarkably lower in all labor outcomes measures. For instance, only 10% of students that would have been rejected by the policy were likely to have a satisfactory performance in the Exit Exam, 74% lower than the probability for the average accepted student; 29% graduated on time (within 6 years after enrollment), 10% lower than the average student accepted; moreover, only 24% of the rejected students started working as teachers after 7 years and only 64% of them worked in good schools, meanwhile the average accepted student were 38% likely to be working after 7 years and 75% of them in good schools; finally, only 12% of the rejected students were classified as good teachers by portfolio examination, half as likely as the accepted teachers.

5.3 Simulating the Screening Rule Back In Time

Figure 17: Screening Rule Back in Time

(a) Rule Applied To Past Teaching Students

(b) Simulation over performance outcomes



Note: Figure 17 shows trends and outcomes for students that would have been admited by P17 from 2007 to 2016. In Figure 17a ■ shows the share of students that would have been rejected by the policy, meanwhile ■ shows the number of students (in thousands) that would have been accepted by the rule. Figure 17a shows the labor outcomes for each group of students enrolled in pedagogy from 2007 - 2016.

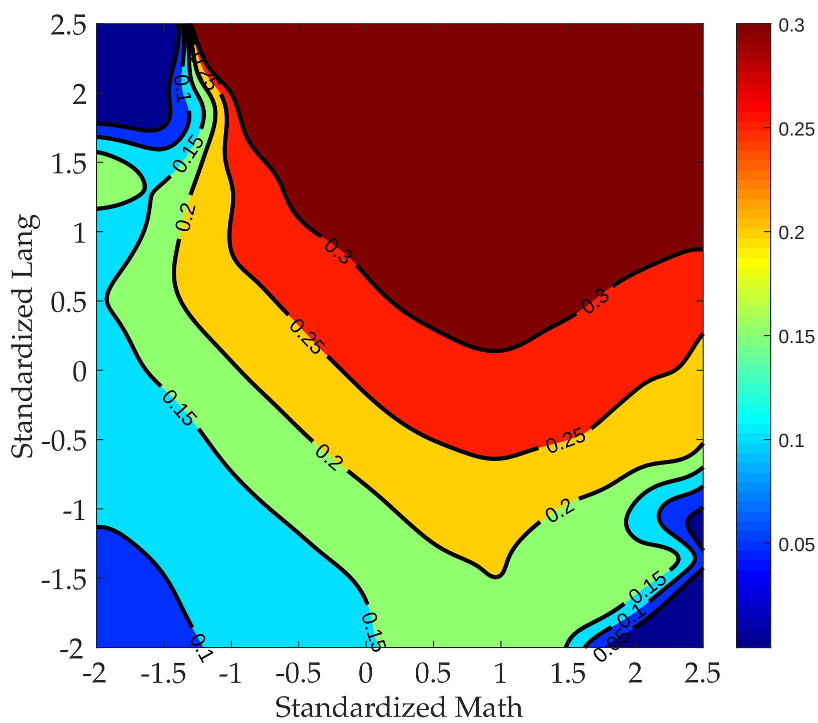
5.4 Towards A More Efficient Screening Policy

The results from the previous section indicate that a simple screening rule can do a fair job ruling out potential unsatisfactory teachers; however, we consider that there is still room to improve the performance of the rule. Increased availability of data and algorithms can help policy-makers make better predictions and design tools for screening talent with greater accuracy. Chile is a special case for exploring such data driven tools thanks to its rich data warehouse which ever year uploads human capital measures and labor outcomes for teachers in the public and private sector. Mixing both algorithms and data can help us guide the design of a more efficient data driven screening policy which screens students that have greater chances of graduating in time and be hired in good schools in the future. Building on evidence presented above, and comparing the performance of the screening policies proposed by the MINEDUC we now try to improve on the simulated outcomes produced by the current policy.

We designed our screening model to be policy invariant by restricting the model to exclusively use input features that would not be affected by outcomes of the policy. In particular, our predictive features are scores on the entrance exam (all of them) and their GPA in their last year of high school. In a second excersize we evaluate how much precision we gave up by not including socioeconomic variables to make the screening rule ethically viable. For this we compare the results of the first model against a model which includes teacher's high school level scores in language and mathematics as well as the average level of education of the parents and average income of the families. Our main outcome variable is being hired by a high school classified as good performing according to SIMCE in less than 7 years of being enrolled in a teaching college.

Figure 18 highlights the potential role for a flexible screening rule, and shows the conditional mean off being hired by a school with high value added as classified by SIMCE 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 rather nonlinear with areas of low probability within the upper left and lower right corners. 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 18: Contour Plot of the Estimated Conditional Mean of Observational Teacher Evaluations



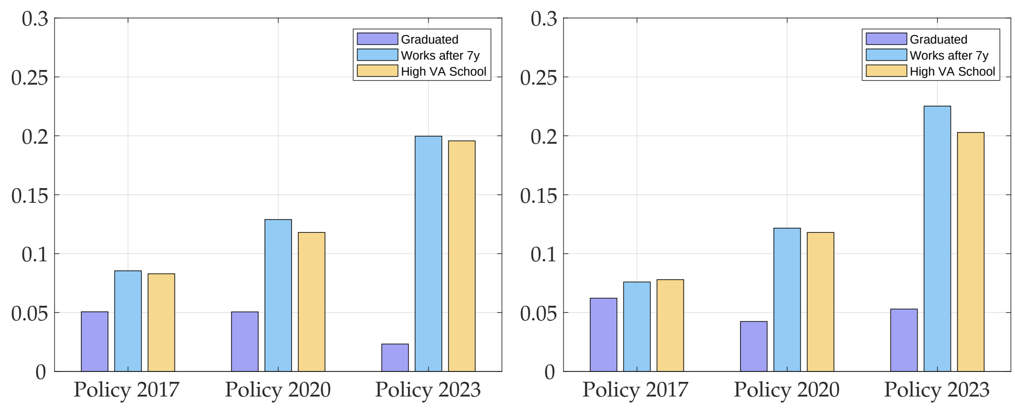
Note: This figure shows the contour plot of the mean of teacher evaluations conditional on pre-college academic achivement in math and language scores.

We estimate and validate our data driven screening strategy by selecting a sample of teachers that allowed us to maximize the number of observations while keeping the maximum number of variables that had a meaningful contribution to the prediction success. In specific, we kept the sample of students that enrolled into a teaching program from 2007 to 2011 primarily because the teachers' registry is only available until year 2018 meaning we could only match the dependent variable with students enrolled until 2011, 7 years before they start working. Additionally, we used this sample because we could not match observations for the years before 2007 with high school level variables of SIMCE. Finally, we dropped all observations without individual math, verbal and NEM scores available in the sample. The resulting sample consists of 36990 teachers enrolled in teaching programs, which was splited in a training set (85%) and test set (15%) consisting of 24778 and 2754 observations respectively.

After crossvalidating different models, we selected the Gradient Boosting Machine which reported an area under the curve of 66% for the SES model and 65% for the model without SES variables. AUC estimates of our crossvalidation across different models, samples and outcomes are in the Online Appendix. Our AUC estimates are higher than the standard for predicting behavioral outcomes **?**.

Figure 19: Outcomes for Those Screened In Simulation ML

(a) Student Level Characteristics **(b)** With school level characteristics



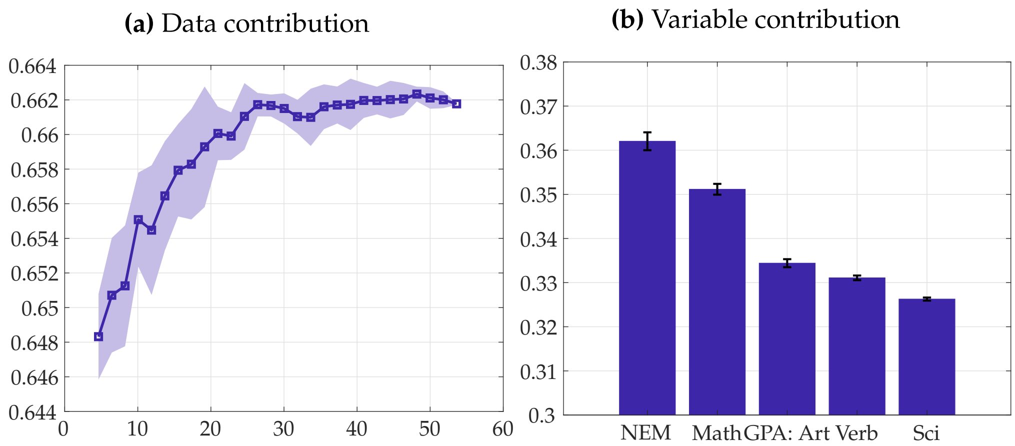
Note: The figures above show the percentage increase in each Graduation, working after 7 years and working in a good school for the students that would have been admited by an ML screening method with a count of students rejected equivalente to those screened out by the rules proposed by the government. Figure 19b shows the results when using student level characteristics and school level characteristics such as socioeconomic status. Figure 19a shows results from the ML model by only using pre college human capital measures such as NEM, scores in PSU exams.

To evaluate the performance of our ML policy as oposed to the government rule, we imposed a probability cutoff for the ML policy that rejects the same number of students that each of the government policies would have rejected in the applicants sample of 2011-2016. The rationale of mantaining the number of students admited constant is to evaluate if the ML screening method has the potential to make less mistakes at (1) admiting prospective satisfactory teachers that would have been mistakenly rejected by (2) rejecting admision to potential bad performers that would have been accepted by the government rule.

Our results suggest that the model does a good job to improve the government rule by only relying on pre college human capital characteristics. Figure 19a shows that although the data driven method does not outperform P17 rule by much (an increase in performance of around 7%), it significantly improves the outcomes of P20 and P23 which are more strict and rejects acceptance to more students. When compared to P20, the data driven method increases the number of students graduating in time in 5%, the number of students working after 7 years of enrollment in 12% and the number of teachers in schools with high value added in 11%. When compared to P23 the improvement is even higher, the ml policy increases the number of timely graduates in 5% and the number of students working at schools and in high value added schools after 7 years in 20%. Our estimates of performance increase are robust to changes in sample size and use of different features as can be seen in the Online Appendix.

In Figure Figure 19b we studied how much performance we give up by not using socioeconomic variables. Results of the exercise suggest that the second ML screening rule performs almost equal to the first method when compared to P17 and P20 and is better than P23 only to predict the share of students working after 6 years (an increase in 3 percentage points). Those results imply that socioeconomic varaibles can contribute to predict success for teachers but the of those variables are not highly significant.

Figure 20: Model accuracy composition



Note: Figure 20a plots the area under the curve evaluated in the test sample obtained by training the same model with different sample sizes (in thousands) as shown in the X axis, the error bars are the cross validation standard errors. Figure 20b shows the prediction loss $1 - AUC$ in terms if we remove independently each of the variables from the model.

Although our model outperforms all government's screening rules, it has some limitations. In Figure 20, we examine up to what extent is it possible to continue increasing the accuracy of our model. For this, we ploted the area under the curve evaluated in the test sample for models trained with different levels of sample size in the training sample. Figure 20a shows that there is still room for improvement in terms of accuracy if the model includes more data to train parameters; however, the contribution seems to have a concave pattern and could be reaching a limit soon. Also, we plotted the contribution of each variable over performance of the model, Figure 20b shows that the most important variable for predicting teaching performance are first, GPA for students in high school and mathematics score in exit exam which are measures of cognitive skills, and third GPA in art during high school which may be an indicator of soft skills.

5.5 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 seems 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 are becoming more abundant as administrative sources become available; historical information gets digitalized and new information gets recorded in more detail than ever before. Combined with the development of improved algorithms, these data are lowering the cost of making increasingly accurate predictions and are influencing decisions, such as hiring, in many markets (Agrawal et al., 2018). In this paper, we put together historical datasets, administrative records and data sources collected in the field to show that better data and flexible prediction methods can be used to implement enhanced teacher screening policies.

In particular, we first 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 students into teacher colleges 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 in most countries (Neal, 2011) 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 while identifying very effective teachers remains challenging, very simple models can use the data to categorize and identify a significant proportion of the worst performing teachers. We find a concave relationship between pre-college academic achievement and later teacher productivity, which we interpret as evidence that in a developing country context such as Chile, basic academic competency is a necessary condition to be an effective teacher. We provide 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 on exit exam scores once conditioning on pre-college academic achievement. Taken together, our evidence seems to suggest that neither training nor experience are enough to undo the significant initial deficiencies that very low performing students have and thus are 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.

The first policy, implemented in 2011, aimed at bringing more qualified students into teacher training programs via financial incentives. At the same time, teacher colleges willing to participate in the policy had to limit admission to students with above average college entrance exams. We show evidence that financial incentives such as scholarships and stipends do indeed lead to higher recruitment of talented students entering participating teacher colleges. The probability that a student with 1σ above the mean would choose to enroll at teacher colleges rose by 40% , over a relatively low base of 10%. Moreover, early productivity indicators measured seven years later, show that those talented students have indeed higher graduation rates, exit exams and employment probabilities, as predicted by their higher college entrance exam scores. This piece of evidence suggest that the relationship between pre-college academic ability and later outcomes is invariant to these types of policies and lends credence to policies using college entrance exam scores as predictors of future performance.

We also show that about half of the teacher colleges decided to participate, which significantly reduced the amount of low performing students matriculating in teacher colleges nationwide. We estimate that screening restrictions decreased the bottom tail of the distribution by one fifth of the total freshmen enrollment (4,000 over 20,000 students). In addition, many higher education options became tuition free as part of another government policy years later (2016). This new policy changed relative prices and generated suggestive evidence helping to disentangle effects attributed to the components described above. In practice, we find that the effectiveness of the financial incentives was significantly reduced. The results suggest that inducing colleges to voluntarily 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 all teaching colleges from admitting students with below average scores unless they had a 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 been 10% more successful than the screening method proposed by the government by using only pre-college human capital characteristics. 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 most effective aspect of the policy comes with screening policies aimed at excluding prospective students with scores below the median rather than 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 it is difficult to recruit high ability students. 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 the 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 US. Exit exams in Chile indicate that 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 Figure 6 and these results are mostly relevant for developing countries.

The policy relevance of screening policies are 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 for them. 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 are less likely to benefit from those studies. Screening policies may seem less relevant for developing countries that are trying to expand educational access and need more teachers and more classrooms, but our findings can be informative for these countries. The temptation is to expand rapidly and lower standards to fill classrooms with bodies. However, recruiting 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.

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 seems 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 contexts and consider the objectives and priorities of the policy-maker.

References

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

Athey, Susan, "The Impact of Machine Learning on Economics," *The Economics of Artificial Intelligence: An Agenda*, May 2019, *Chapter 21*, 507–547.

— **and Guido W. Imbens**, "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 2019, *11* (1), 685–725.

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.

Biasi, Barbara, "The Labor Market for Teachers Under Different Pay Schemes," NBER Working Papers 24813, National Bureau of Economic Research, Inc July 2018.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson, and Waly Wane, "Enrollment without Learning: Teacher Effort, Knowledge, and Skill in Primary Schools in Africa," *Journal of Economic Perspectives*, November 2017, *31* (4), 185–204.

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

Castro-Zarzur, Rosa and Carolina Mendez, "Effectiveness of the BVP in Time," Working Paper, Mimeo. 2019.

— **, Miguel Sarzosa, and Ricardo Espinoza**, "Unintended Consequences of Free College: Self-Selection into the Teaching Profession," Working Paper, Mimeo. 2019.

Chalfin, Aaron, Oren Danieli, Andrew Hillis, Zubin Jelveh, Michael Luca, Jens Ludwig, and Sendhil Mullainathan, "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review P&P*, May 2016, *106* (5), 124–27.

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.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff, "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood," *American Economic Review*, September 2014, *104* (9), 2633–79.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor, "Teacher credentials and student achievement: Longitudinal analysis with student fixed effects," *Economics of Education Review*, 2007, *26* (6), 673 – 682. Economics of Education: Major Contributions and Future Directions - The Dijon Papers.

Coffman, Lucas C., Clayton R. Featherstone, and Judd B. Kessler, "Can Social Information Affect What Job You Choose and Keep?," *American Economic Journal: Applied Economics*, January 2017, *9* (1), 96–117.

Coffman, Lucas C, John J Conlon, Clayton R Featherstone, and Judd B Kessler, "Liquidity Affects Job Choice: Evidence from Teach for America*," *The Quarterly Journal of Economics*, 06 2019, *134* (4), 2203–2236.

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.

de Ree, Joppe, Karthik Muralidharan, Menno Pradhan, and Halsey Rogers, "Double for Nothing? Experimental Evidence on an Unconditional Teacher Salary Increase in Indonesia*," *The Quarterly Journal of Economics*, 11 2017, *133* (2), 993–1039.

Elacqua, Gregory, Diana Hincapie, Isabel Hincapie, and Veronica Montalva, "Can Financial Incentives Help Disadvantaged Schools to Attract and Retain High-performing Teachers?: Evidence from Chile," IDB Working Paper Series 1080, Inter-American Development Bank November 2019.

Estrada, Ricardo, "Rules versus Discretion in Public Service: Teacher Hiring in Mexico," *Journal of Labor Economics*, 2019, *37* (2), 545–579.

Figlio, David, Krzysztof Karbownik, and Kjell Salvanes, "The Promise of Administrative Data in Education Research," *Education Finance and Policy*, 2017, *12* (2), 129–136.

Figlio, David N. and Lawrence W. Kenny, "Individual teacher incentives and student performance," *Journal of Public Economics*, 2007, *91* (5), 901 – 914.

Gilligan, Daniel O, Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M Lucas, and Derek Neal, "Educator Incentives and Educational Triage in Rural Primary Schools," Working Paper 24911, National Bureau of Economic Research August 2018.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics*, July 2010, *2* (3), 205–27.

Gregorio, Soledad De, Sebastian Gallegos, and Christopher Neilson, "Decomposing Teachers Contributions to Learning," Working Paper, Mimeo 2019.

Hanushek, Eric, "The Economic Value of Higher Teacher Quality," *Economics of Education Review*, 2011, *30* (3), 466–479.

Hanushek, Eric A., Marc Piopiunik, and Simon Wiederhold, "The Value of Smarter Teachers: International Evidence on Teacher Cognitive Skills and Student Performance," *Journal of Human Resources*, Fall 2019, *54* (4), 857–899.

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

Harris, D. and T. Sass, "Teacher training, teacher quality, and student achievement," *Journal of Public Economics*, 2011, *95*, 798–812.

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.

Hoxby, Caroline Minter, "How Teachers' Unions Affect Education Production*," *The Quarterly Journal of Economics*, 08 1996, *111* (3), 671–718.

Jackson, C Kirabo, "Recruiting, retaining, and creating quality teachers," *Nordic Economic Policy Review*, 2012, *3* (1), 61–104.

Jackson, C. Kirabo, Jonah E. Rockoff, and Douglas O. Staiger, "Teacher Effects and Teacher-Related Policies," *Annual Review of Economics*, August 2014, *6* (1), 801–825.

Jacob, Brian A., "Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools," *Journal of Public Economics*, 2005, *89* (5), 761 – 796.

— **and Lars Lefgren**, "The Impact of Teacher Training on Student Achievement: Quasi-Experimental Evidence from School Reform Efforts in Chicago," *The Journal of Human Resources*, 2004, *39* (1), 50–79.

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

Kane, Thomas J., Jonah E. Rockoff, and Douglas O. Staiger, "What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City," *Economics of Education Review*, 2008, *27* (6), 615–631.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, "Human Decisions and Machine Predictions*," *The Quarterly Journal of Economics*, 08 2017, *133* (1), 237–293.

Lombardi, María, "Is the remedy worse than the disease? The impact of teacher remediation on teacher and student performance in Chile," *Economics of Education Review*, 2019, *73*, 101928.

Majerowicz, Stephanie and Ricardo Montero, "Can Teaching be Taught? Experimental Evidence from a Teacher Coaching Program in Peru (Job Market Paper, Harvard Kennedy School)," 2018.

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.

Mbiti, Isaac, Mauricio Romero, and Youdi Schipper, "Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania," NBER Working Papers 25903, National Bureau of Economic Research, Inc May 2019.

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

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, May 2017, *31* (2), 87–106.

Muralidharan, Karthik and Venkatesh Sundararaman, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011, *119* (1), 39–77.

Neal, Derek, "Chapter 6 - The Design of Performance Pay in Education," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of The Economics of Education*, Vol. 4 of *Handbook of the Economics of Education*, Elsevier, 2011, pp. 495 – 550.

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

OECD, *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, Paris: Organisation for Economic Co-operation and Development, 2005.

_ , *Education at a Glance 2009: OECD Indicators*, Paris: Organisation for Economic Co-operation 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.

Rothstein, Jesse M., "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions," *American Economic Review*, September 2006, *96* (4), 1333–1350.

Sajjadiani, S., A. J. Sojourner, J. D. Kammeyer-Mueller, and E. Mykerezi, "Using machine learning to translate applicant work history into predictors of performance and turnover," *Journal of Applied Psychology*, 2019, *104* (10), 1207–1225.

Svensson, Tessa Bold Deon P. Filmer Ezequiel Molina Jakob, *The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa*, The World Bank, 2019.

Tincani, Michaela, "Teacher Labor Markets, School Vouchers and Student Cognitive Achievement: Evidence from Chile," *Under revision, Quantitative Economics*, 2018.

__ , Petra Todd, Jere R. Behrman, and Kenneth I. Wolpin, "Teacher Quality in Public and Private Schools Under a Voucher System: The Case of Chile," *Journal of Labor Economics*, 2016, *34* (2).

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 4: 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 5: 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 6: 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 7: 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 8: 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 9: 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 10: 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	

Table 11: Summary statistics for ML train and test set

	Mean Train	Mean Test	Difference
Outcome Variable			
Good School	0.25	0.25	0.00
Student Level			
NEM	539.47	539.99	-0.53
PSU Verbal	522.50	522.86	-0.36
PSU Math	506.65	507.68	-1.03
Didn't took History	0.23	0.24	-0.01
Didn't took Science	0.56	0.56	0.01
School Level			
SIMCE MATH	268.00	268.47	-0.46
SIMCE LANG	276.62	277.04	-0.42
SIMCE MATH STD	44.22	44.21	0.01
SIMCE LANG STD	44.87	44.82	0.05
Mother Level 1	0.14	0.14	0.00
Mother Level 2	0.25	0.25	0.00
Mother Level 3	0.39	0.39	0.00
Mother Level 4	0.15	0.14	0.00
Mother Level 5	0.06	0.06	0.00
Father Level 1	0.18	0.18	0.00
Father Level 2	0.21	0.21	0.00
Father Level 3	0.35	0.35	0.00
Father Level 4	0.15	0.15	0.00
Father Level 5	0.06	0.06	0.00
Income SIMCE	62.32	62.47	-0.16
Income SIMCE STD	19.71	19.73	-0.02
N Observations	29592	7398	

Note: All the differences in means are not statistically significant meaning that the means in the train and test sample are statistically the same.

Appendix

Timeline of Higher Education Reforms in Chile[14]

1967 Implementation of the college entrance exam PAA

1981 Systemic reform: Funding shifted from government to students; New types of institutions were created; and entry barriers to the marker were lowered.

1982 Creation of the National Fund for Scientific and Technological Research (FONDECYT). This is a research fund to be granted under a competitive system with external experts that evaluate the proposals.

1990 Creation of Council of Education (Consejo Superior de Educacion): This council was created to be the organization responsible for managing the accreditation system created by the LOCE (Ley Organica Constitutional de Educacion).

1994 Funding to give more access to incoming students was increased, adding several instruments to the student funding scheme.

- Several scholarships were created: Juan Gomez Millas; High-performing Students in the Teaching Profession; Children of Education Professionals; and Work Performance for Higher Education students
- The Solidary University Credit Fund was put in place, with flexible payments according to income and an unified system of socioeconomic assessment (Formulario unico de Acreditacion Socioeconomica, FUAS) to improve the allocation of financial aid.
- Student Loans: in 1995, the Production Development Corporation (Corporacion de Fomento de la Produccion, CORFO) created a special loan to finance graduate studies. In 1997, a similar loan for undergrad studies was created.

1996 The Institutional Development Fund was created. It focused in strengthening regional universities, and in promoting innovation in undergrad teaching.

1998 29 Performance Agreement for Development of Priority Areas (Convenios de Desempeno para el Desarrollo de areas Prioritarias) were made with 21 institutions. This program helped as a pilot program for the Higher Education Improvement Program: (Mejoramiento de la Equidad y Calidad de la Educacion Superior, or MECESUP)

1998 Creation the Higher Education Improvement Program (Mejoramiento de la Equidad y Calidad de la Educacion Superior, or MECESUP). The objective of this program is to help higher education institutions to improve the quality of their programs.

1998 Juan Gomez Millas. High-performing Students in the teaching Profession scholarships

1999 Creation of the CNAP (Undergraduate National Accreditation Commission)

2003 Implementation of the PSU

2005 Creation of the Government Guaranteed Loan (CAE) managed in partnership with commercial banks, open to students in CRUCH or accredited non-CRUCH higher education institutions.

[14] Maria Fernanda Ramirez Espinoza contributed with most of the information provided in this section.

2006 Creation of the National System of Quality Assurance: This program involves the accreditation of institutions and study programs. Accreditation is voluntary in the sense that institutions may continue to operate without it; but certain types of student support are available only to students at accredited universities, and certain programs (such as teaching and medicine) must be accredited if they are to be publicly funded. The 2006 law built on the practice and procedures developed under the former fully voluntary accreditation system originating in the 1990s.

2006 PSU scholarship. It is aimed to all students that cannot pay to register for the university entry exams.

2007 Creation of the Reference Tuition Fees: Reference tuitions are used to calculate the maximum amount of student aid (grant and loan) that eligible students (based on income criteria) are eligible for. This system classifies institutions based on an index that considers academic degrees, approved research projects, publications, graduation rates and retention rates of first year to determine the annual reference tuition for each program in each institution.

2011 BVP

2012 The interest rate of the CAE credit is reduced from about 5.8% to 2%. In addition, the monthly payments of the credit are capped at 10% of the income.

2015 Through the national budget, students that come from the 50% lower income have a scholarship if they get accepted to institutions that comply with certain requisites.

2017 Free College