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 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, we first document a robust positive and concave relationship between pre-college academic achievement and several short and long run measures of teacher productivity. We then assess the effectiveness of two recent policies that used pre-college achievement to recruit or screen out students entering teacher-colleges. Using a regression discontinuity design based on the government's recruitment efforts, we evaluate the effectiveness of targeted scholarships at shifting career choices of high achieving students as well as the effect on the overall stock of teachers predicted effectiveness. We then assess a screening policy that forced teacher colleges to exclude below-average applicants. We quantify the policy effectiveness by retroactively simulating the rule and evaluating its success at screening out low performing teachers. Comparing this benchmark policy rule to a series of data-driven alternatives, we find that even simple screening policies can identify a significant portion of ex-post low performing teachers. In both policies studied, screening out 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 their teachers (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, which affect the stock of teachers in the short run. In the medium and long term, recruiting and screening policies can be convenient compared to the on-the-job policies for several reasons. The first is that successful recruiting policies can prevent students from exposure to ineffective teachers, who are usually difficult to remove once employed. Second, it is logistically and politically hard to implement pay for performance schemes that look to incentivize effort (Hoxby, 1996; Hanushek, 2011). Third, the evidence suggests that later investments in training have a little influence on teacher productivity (Jackson, 2012; Lombardi, 2019). Targeting investments attracting individuals who have the highest chance of being effective teachers 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 scarce evidence that the data that does exist can reliably predict teacher effectiveness before entering higher education.

Data is now becoming more abundant than ever before. Administrative sources and historical records are being digitalized and governments are developing the capacity to store and analyze the data (Figlio et al., 2017). Combined with the development of improved algorithms, the cost of making increasingly accurate predictions is lowering and 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 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 achievement 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 regression discontinuity 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 effectiveness 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. 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.

The observed correlation between entrance exams and later outcomes could be 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. Using 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 result suggest that college training is not enough to undo initial differences and that pre-college academic readiness has a 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 policy increased the number of higher scoring students in teacher colleges, with the highest effects at the lower cutoffs of the college entrance distribution (about 37% of an effect size). Eight years later, we find that the higher-scoring students went on to work in schools later on (effect size of 34% on employment at schools). This finding indicates that the policy was successful at raising the predicted quality of students who entered into the teaching profession.[1]

We also measured other early indicators of such as graduation rates and the exit exams, finding precise zero effects. These results suggest that the higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores, indicating that the predicted relationship between pre-college academic achievement and teacher medium run outcomes is 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

[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).

to admit applicants with scores below the national mean. Using another RD at the national mean cutoff we estimate that the policy reduced the fraction of low-scoring students enrolled in teacher colleges by half in 2017 and 2018.

We replicate the policy rule back in time to describe who would have been affected and whether the excluded students become ineffective or effective 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.

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 entrance exams and high school transcript data. 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 6%, and increased the number of teachers working in well performing schools in 6%.

These results are important because they have direct policy implications. If teacher effectiveness, or lack thereof, is possible to predict early on, then 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 extremely low levels of prior academic achievement.

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 envi-

[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.

ronment 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 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.

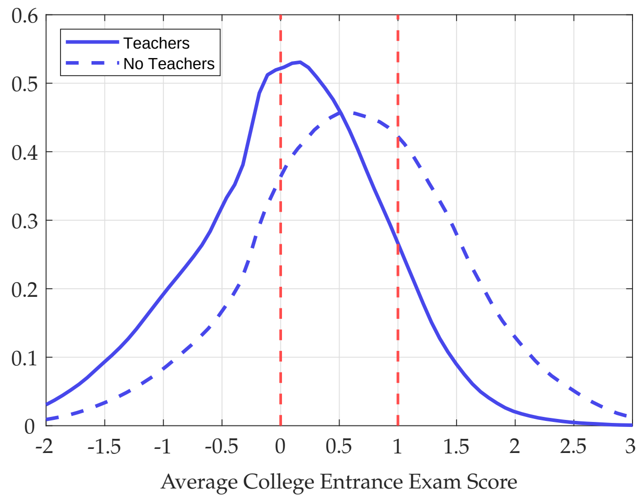
Attracting more skilled individuals to be teachers is challenging because, among other factors, teachers are typically paid less than comparable professionals (Mizala and Nopo, 2016).[4] Consistently, we know from the related literature that college graduates with higher college entrance scores are less 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



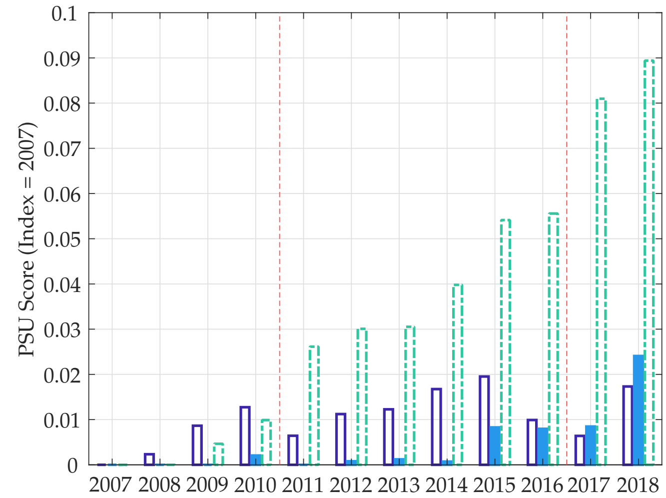
Note: Figure 1 plots the distribution of college entrance exam scores for freshmen in teacher colleges (continuous line) and freshmen in the health and STEM fields (dotted line), using data from 2007 to 2010. The entrance exam score (in standard deviation units) is the average of the math and language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



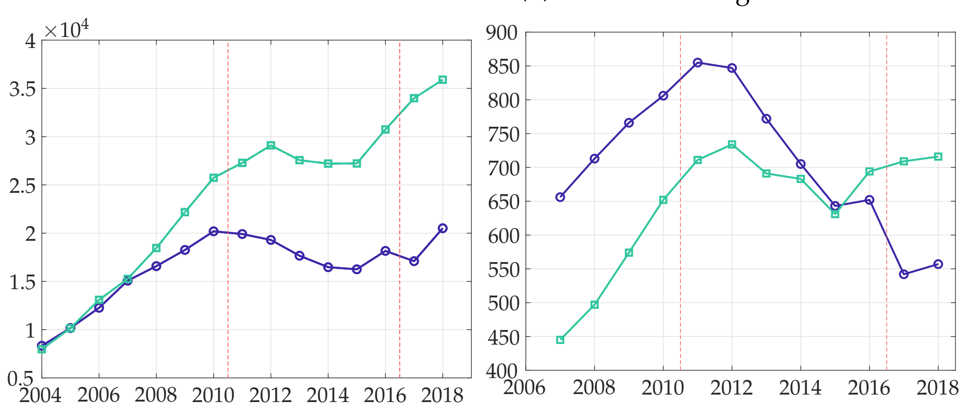
Note: In Figure 2 the dotted bar plots the percent change on the average college entrance exam score (labeled 'PSU Score') for freshmen in teacher colleges, the filled bar does the same for health, and the empty bar for STEM careers, from 2007 to 2018, using 2007 as the base year. The dotted vertical 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]

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.

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

[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.

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.

3 Pre-college Achievement and Teacher Productivity

In this section we document the systematic correlation between pre-college academic ability and the teacher productivity measures described in the previous 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.

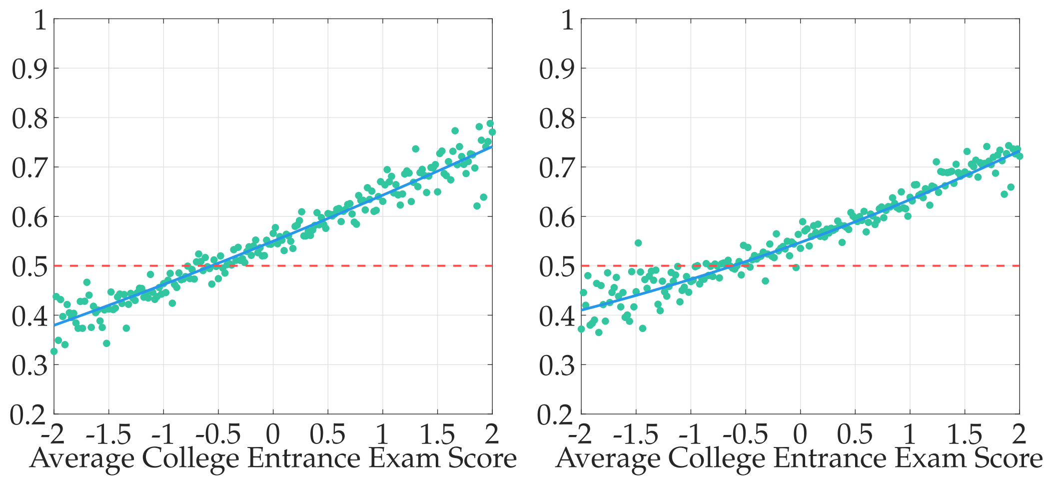
The general takeaway is that the empirical relationship between pre-college skills as a student and teacher productivity later on is positive and concave. In Table 1 we show the coefficients of 14 separate regressions for different measures of teacher performance on the college entrance exam score (in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

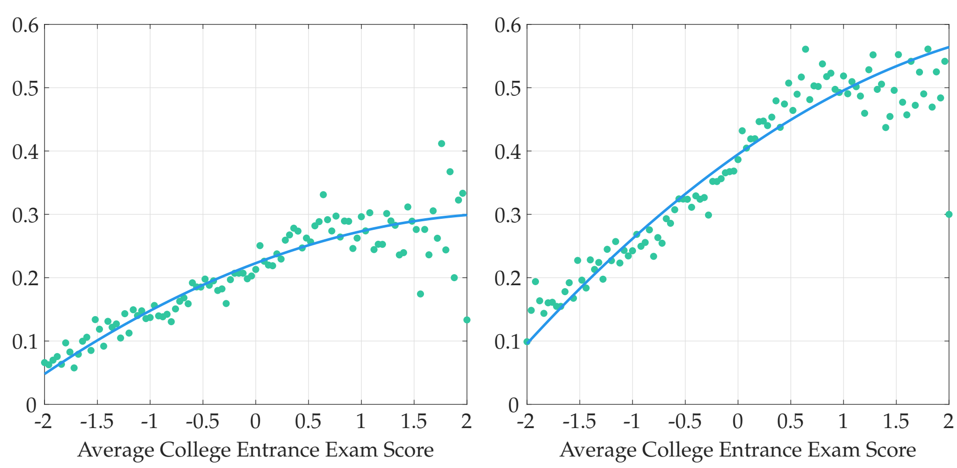
(a) Disciplinary Exit Exam **(b)** General Pedagogy Exit Exam



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

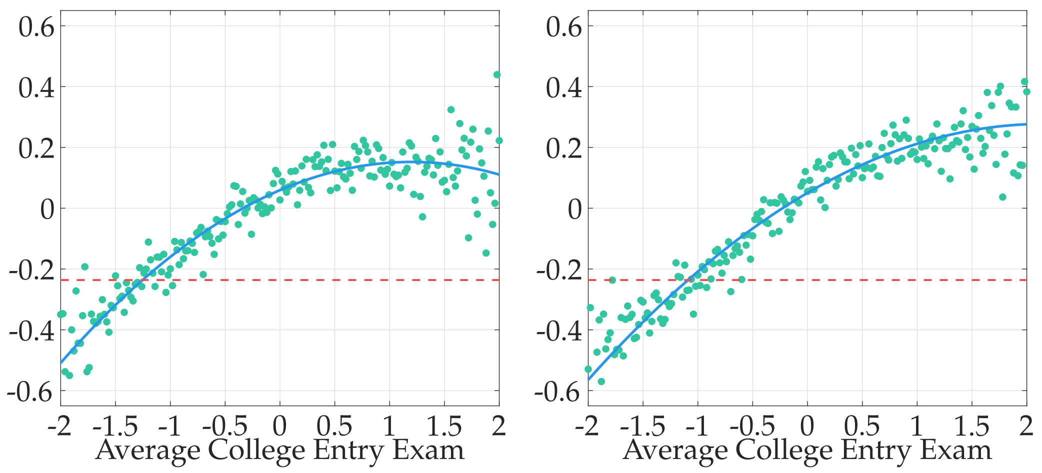


Note: The figures plot the probability of graduation after 5 years (Figure 5a) and 8 years (Figure 5b) of first enrollment, within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in students enrolled in years 2004 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



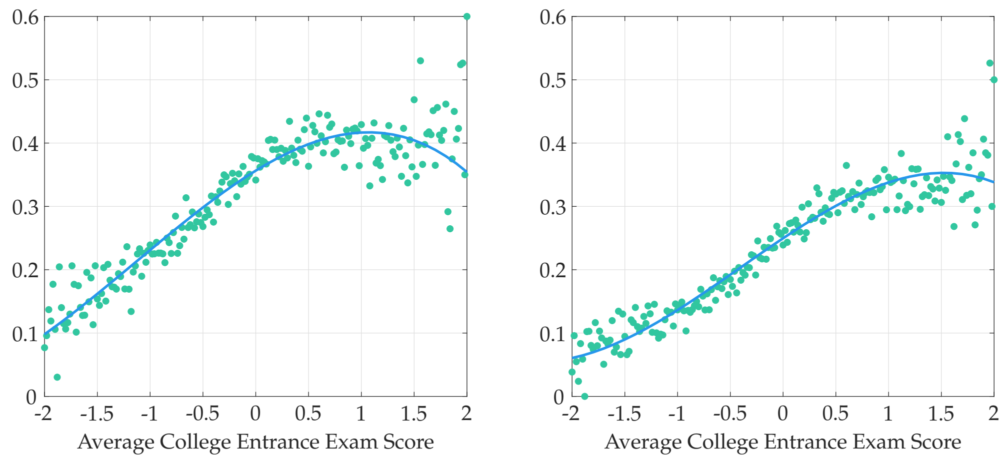
Note: The figures plot the teacher evaluation scores (overall in Figure 6a and the portfolio component in Figure 6b), within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in teachers evaluated between years 2004 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

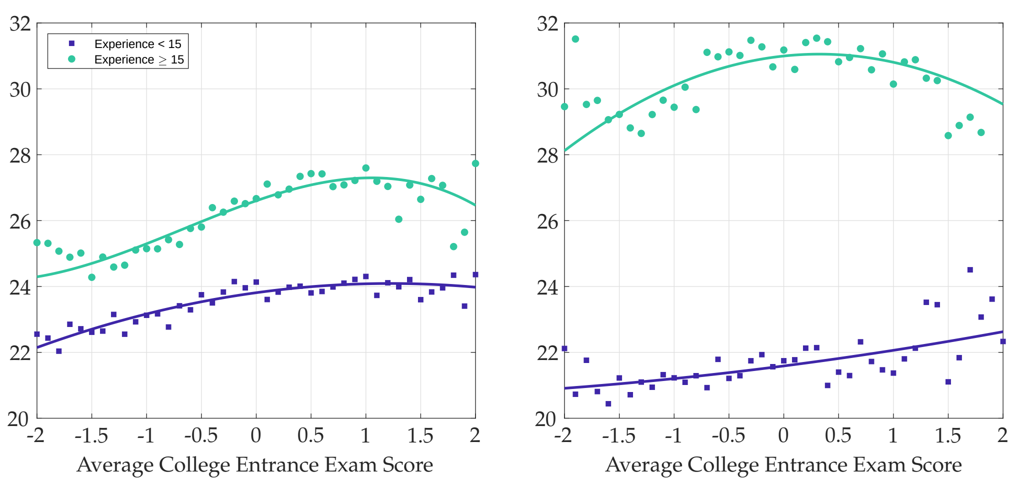
(a) Employment in Schools **(b)** Employment in High VA Schools



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)

(a) Private Sector **(b)** Public Sector



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)		
$(\text{PSU Score})^2$	-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)
$(\text{PSU Score})^2$	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)
$(\text{PSU Score})^2$	-0.048 ***	-0.031 ***	-0.049 ***	-0.055 ***
	(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**	
PSU Score	0.298***	0.260***	0.269***	
	(0.044)	(0.044)	(0.089)	
$(\text{PSU Score})^2$	-0.027***	-0.025***	-0.024***	
	(0.114)	(0.113)	(0.235)	
Observations	[13,201]	[13,201]	[13,201]	
Dep. Var. Mean	0.470	0.435	0.287	

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.

Taken together, these correlations suggest that pre-college achievement could be used to predict teacher quality later on. We study whether access to higher value added teacher colleges causes these observed correlations. We combine a regression discontinuity design with data on the population of applicants to teaching colleges from 1977 to 2011, to estimate the value added of teaching colleges versus the next best alternative. The findings which we present in our Online Appendix, show no evidence that any particular teaching college adds more value or contributes to closing or increasing the predicted gap in teacher effectiveness.

This result suggests that pre-college achievement could be useful for policies aimed to recruit or screen out students entering teacher-colleges. In the next sections of the paper we study two recent policy implementations that look to screen out low performing students from teachers colleges or to attract high achieving students to teachers colleges.

4 A *Carrot & 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%, while enrollment from the lower end of the pre-college achievement 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 policy, first implemented in 2011, offered full scholarships and other incentives such as stipends and paid semesters abroad for high scoring test-takers who enroll as freshmen at teacher colleges. One distinguishing characteristic of this scholarship is that it had no socioeconomic requirement.

To be eligible test-takers should have achieved scores from approximately the highest 20% of the college entrance exam distribution.[8]

In particular, students with scores in the top 20% (i.e., 600 points or more) were eligible for a full tuition scholarship.[9] If they scored above $\mu + 2\sigma$ (700 points, top 5%), they were eligible for the full tuition scholarship plus a monthly stipend of about $US150 (close to 50% of the minimum wage). With scores above $\mu + 2.2\sigma$ (720 points, top 2%), enrollees would benefit from tuition, stipend and 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 participating teacher colleges to screen out low scoring applicants. In particular, colleges were required to implement a minimum cutoff score at the national mean of 500 points if they wanted their students to benefit from the BVP.[10] In addition, participating teacher colleges needed to be accredited for at least 2 years at all campuses as determined by the National Commission of Accreditation (CNA).

4.2 Empirical Strategy and Data

We use a regression discontinuity (RD) exploiting the BVP score cutoffs to evaluate whether the policy attracted higher-scoring test-takers to teacher colleges.

Our identifying assumptions are standard for RD designs. We assume that there are no other changes occurring at the thresholds that could confound our estimates. In our Appendix we run a series of robustness tests showing that there are no differences in a host of covariates around the thresholds, no evidence of score manipulation, and also shows that our estimates are stable to using different bandwidths and specifications.

[8] 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 were not eligible for the scholarship.

[9] If the student had obtained another scholarship called *Beca Excelencia Academica* the cutoff will be 580. These are a handful of students (N=61) and do not change our results if included in the analysis.

[10] The cutoff was lax, allowing colleges to enroll a maximum of 15% of their entering class starting in 2011 with scores below the cutoff.

Our main estimating equation is

$$Y_i = \alpha_0 + \alpha_1 Z_i + f(S_i) + \alpha_2 X_i + \mu_i. \quad (1)$$

where Y_i represents a particular outcome such as enrollment at teacher colleges for test-taker i. Our parameter of interest is α_1, which is the *intention-to-treat* effect of the BVP policy on the outcome Y_i. The indicator variable Z_i is equal to 1 if the test-taker i achieved a score above a particular threshold and zero otherwise. For simplicity, we estimate separate regressions for the 500, 600, 700 and 720 policy cutoffs.[11] $f(S_i)$ is a smooth function of scores that includes interactions with Z_i to allow for different slopes on each side of the cutoff, and μ_i represents the error term that we cluster within the college entrance exam scores. We also include a set of predetermined variables as controls in X_i, such as test-takers' gender, household income, parents education, region of residence, and type of school. In practice, these control variables have very little effect on our RD estimates and serve mainly to improve precision.

We implement our empirical strategy using individual level data from the college entrance exam. We first present results for the 2011 cohort, for whom we can estimate the immediate take-up and enrollment effects, but also later outcomes like graduation, exit exams and employment in schools up to 2019. We also compute short-run estimates for later cohorts in the following section.

In Table 2 we show descriptive statistics for all test-takers in 2011, organized by information on scores, demographics, and higher education enrollment.[12] The scores have a mean of about 500 points each. A total of 250,758 students took the college entrance exam in December 2010, aiming to start classes when the academic year starts, in March 2011. All of these test-takers were potentially eligible for the BVP if they achieve scores above the policy cutoffs.

Test takers are on average 19 years old at the moment of the test, and about half of them are girls. Their parents have on average slightly more than 11 years of completed schooling, and about 40% lives in the capital city. All this figures are consistent with data coming from national surveys (CASEN 2016) and censuses. About 55%, 35% and 10% graduated from voucher, public and private high schools, which again are consistent with population figures on enrollment in the country (MINEDUC 2018).

The last panel shows the fraction of test takers who enroll in higher education. A 63% of them enroll at any institution, 44% enrolls at colleges and half of that enrolls at the CRUCH universities. An 8% enrolls at any teacher college and a 5% enrolls at teacher colleges that were BVP eligible.

[11] We also ran a more complex version of Equation 1 to estimate all threshold effects jointly with no differences in our results.

[12] Test-takers complete a survey providing information on their gender, date of birth, household income bracket and parental schooling among other characteristics. We combine this data with the scores information at the individual level, which we merge with administrative records of higher education enrollment coming from the MINEDUC. The enrollment records have information for the population of students enrolled in higher education institutions in the country.

Table 2: Descriptive Statistics for all Test-Takers

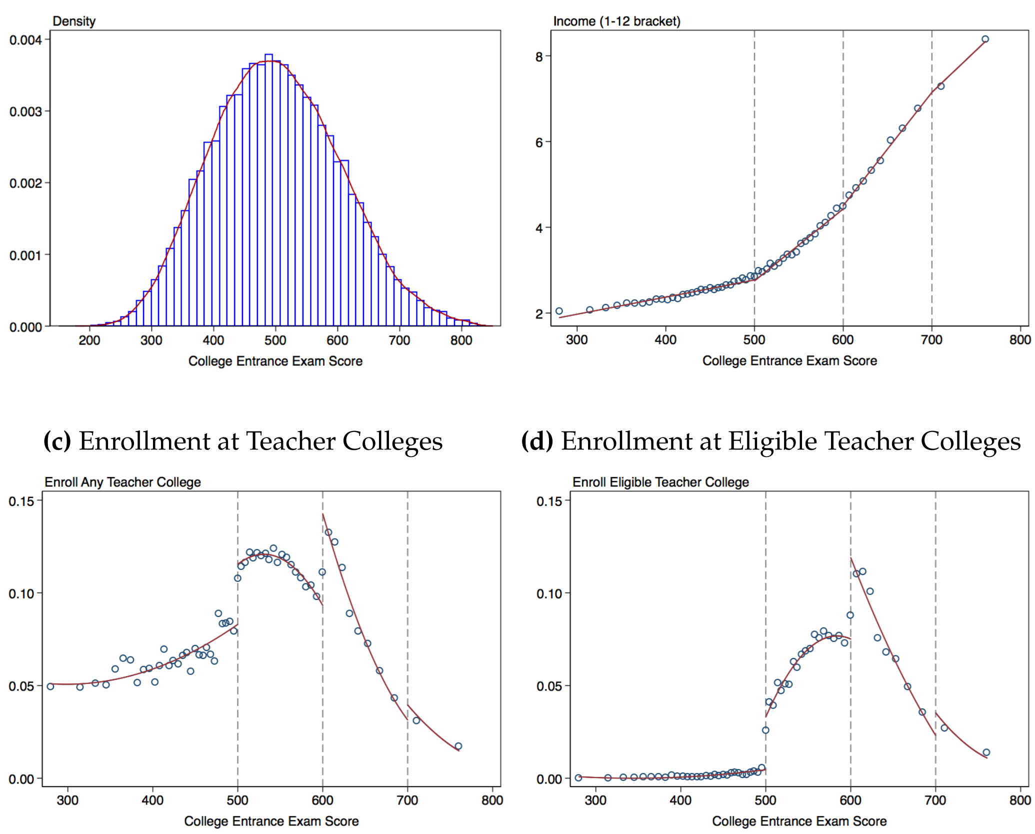
Variable	(1) Observations	(2) Mean	(3) Std. Deviation	(4) Min	(5) Max
Scores					
College Exam Score	250,758	501.06	102.34	178	850
Math Score	250,758	501.07	111.27	150	850
Language Score	250,758	501.04	108.34	150	850
Takes History Test	250,758	0.62	0.49	0	1
History Score	154,790	500.41	109.55	150	850
Takes Science Test	250,758	0.56	0.50	0	1
Science Score	139,783	500.52	109.47	150	850
High School GPA Score	248,807	535.81	99.88	208	826
Demographics					
Female	250,758	0.52	0.50	0	1
Age	250,758	19.38	3.17	15	78
Income (1-12 bracket)	250,758	3.40	2.88	1	12
Private Health Insurance	250,758	0.21	0.40	0	1
Father Schooling (years)	215,105	11.45	3.77	0	17
Mother Schooling (years)	233,044	11.30	3.57	0	17
Capital City	248,462	0.40	0.49	0	1
Public High School	248,462	0.35	0.48	0	1
Private High School	248,462	0.10	0.30	0	1
Voucher High School	248,462	0.55	0.50	0	1
Enrollment					
Enroll Higher Education	250,758	0.63	0.48	0	1
Enroll College	250,758	0.44	0.50	0	1
Enroll CRUCH	250,758	0.21	0.41	0	1
Enroll Any Teacher College	250,758	0.08	0.28	0	1
Enroll Eligible Teacher College	250,758	0.03	0.18	0	1

Notes: Table 2 shows descriptive statistics for the 250,758 students took the college entrance exam in December 2010. The college entrance exam score is the math-language average score; the history and science tests are optional exams. The High School GPA Score has valid data for 99.2% of the test-takers (248,807 of 250,758). The Age corresponds to the age at the moment of the test. The variables of parental schooling have missing information due to both non-response and test-takers not knowing the answer. Capital City indicates whether the test-taker lives in the capital of the country at the moment of the test, while the variables Public, Private and Voucher High School indicate the type of high school from which the test-takers graduated. These last four variables have a response rate of 99.1%. The enrollment variables come from population records collected by the Ministry of Education. Enroll in Higher Education takes value one if the test-taker enrolled at any institute or university. Enroll College is equal to one if the test-taker enrolled at any college; enroll CRUCH does the same if the test taker enrolled at universities belonging to the *Consejo de Rectores*. Enroll at any teacher college (TC) takes value one if test taker enrolled in any education major in the country, and Enroll Eligible TC does the same for enrollment at eligible teacher colleges.

4.3 Results

Our main results show that the policy attracted higher scoring test-takers to teacher colleges. Figure 9 summarizes the first set of findings. Figure 9a and Figure 9b are robustness tests, showing no manipulation of the running variable (the college entrance exam score) and that other covariates, such as household income behave smoothly near the policy thresholds. Figure 9c and Figure 9d illustrate effects on enrollment at any teacher colleges (TC) and at eligible TC, respectively. Both Figures reveal a sharp discontinuity at the 500 and 600 points and a smaller increase at 700 points, indicating that test-takers with very similar scores around those cutoffs experienced a very different likelihood of enrolling at teacher colleges.

Figure 9: Main Results



(a) Score Distribution

(b) Household Income

(c) Enrollment at Teacher Colleges

(d) Enrollment at Eligible Teacher Colleges

Note: Figure 9a plots the distribution of scores for all test takers. Figure 9b, Figure 9c and Figure 9d plot the mean of the y-axis variable within bins of scores, and fit estimated lines using all the underlying data. The sample size in each graph in Figure 9 is of N=250,758 observations.

In Table 3 we show the corresponding point estimates from Equation 1 on teacher college enrollment. The columns show effects at the 500, 600, 700 and 720 cutoffs, with optimal bandwidths for each threshold. These are our preferred estimates, which are robust to different bandwidths and specifications as we show in the Appendix. In addition, effects are zero for years before the policy was implemented, as shown in Figure/Table X.

The estimates from Panel 1 indicate that enrollment at teacher colleges increased by 3.2 percentage points (pp), 3.5 pp., and 2.5pp at the 500, 600 and 700 cutoffs respectively. We find no effect at the highest cutoff of 720 points, precisely estimated. The magnitude of the estimates is sizable for the first three thresholds, representing relative increases of 37% at 500 (3.2pp over 8.6pp just below the cutoff), 37% at 600 (3.5pp over 9.5pp) and 100% at 700 (2.5pp over 2.5pp).

Panel 2 in Table 3 shows similar points estimates for the respective cutoffs on enrollment at eligible teacher colleges. The main difference is that the enrollment rate at eligible teacher colleges just before the cutoff of 500 points is zero, consistent with the policy design. These results suggest the effects are indeed driven by the BVP policy.

Table 3: BVP Effects on Enrollment

Panel 1. Dep. Variable: Enrollment at Teacher Colleges

	(1)	(2)	(3)	(4)
RD_Estimate	0.032***	0.035***	0.025**	-0.010
	(0.004)	(0.007)	(0.009)	(0.008)
Mean Just Below Cutoff	.086	.095	.025	.032
Optimal Bandwidth	48.3	34.3	26.3	34.5
Cutoff Value	500	600	700	720
Effective Observations	86,457	40,559	8,423	8,210
All Observations	250,758	250,758	250,758	250,758

Panel 2. Dep. Variable: Enrollment at Eligible Teacher Colleges

	(1)	(2)	(3)	(4)
RD_Estimate	0.033***	0.029***	0.022**	-0.008
	(0.002)	(0.006)	(0.008)	(0.007)
Mean Just Below Cutoff	.005	.073	.022	.027
Optimal Bandwidth	41.8	30.7	28.4	33.3
Cutoff Value	500	600	700	720
Effective Observations	75,825	36,437	9,178	7,719
All Observations	250,758	250,758	250,758	250,758

Notes: Table 3 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variables are Enrollment at Teacher Colleges and Enrollment at Eligible Teacher Colleges in Panels 1 and 2, respectively. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for high school GPA and the demographics described in Table 2.

4.3.1 *Medium Run Effects*

Our previous results show that the policy attracted higher scoring test takers to enroll at teachers colleges. In this section we examine results on a host of outcomes described in our Correlations Section, like graduation, exit exams and employment in schools, all measured up to eight years after initial enrollment.

A first outcome of interest if whether the policy resulted on the higher scoring test takers actually working at schools later on. Our results in Panel 1 in Table 4 show that the policy increased employment at schools of the higher scoring test-takers near the cutoffs, at the 500 and 600 thresholds. The effect sizes are of 12% at 500 (1.2pp over 6.4pp) and 34% at 600 (2.3 over 6.7).

Panels 2 to 4 show effects on graduation, and the likelihood of taking the exit exam and the teacher evaluation. We find zero effects on these outcomes, with small standard errors. These precise zero effects suggest that higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores.[13]

[13] Panel 5 shows no effects the exit exam score but effects are much noisier due to the low number of observations (2% of the sample took the exit exam overall).

Table 4: BVP Effects on Medium Run Outcomes (8 years)

Panel 1. Dep. Variable: Employment at Schools

	(1)	(2)	(3)	(4)
RD_Estimate	0.012***	0.023***	0.006	-0.010
	(0.003)	(0.005)	(0.008)	(0.007)
Mean Just Below Cutoff	.064	.067	.033	.029
Optimal Bandwidth	60.7	52.5	32.1	38.3
Cutoff Value	500	600	700	720
Effective Observations	107,517	62,410	10,612	9,042
All Observations	250,758	250,758	250,758	250,758

Panel 2. Dep. Variable: Graduation

	(1)	(2)	(3)	(4)
RD_Estimate	-0.000	0.003	-0.013	0.019
	(0.006)	(0.008)	(0.021)	(0.020)
Mean Just Below Cutoff	.522	.575	.600	.626
Optimal Bandwidth	63.9	54	31.8	43.6
Cutoff Value	500	600	700	720
Effective Observations	112,474	63,569	10,328	10,523
All Observations	250,758	250,758	250,758	250,758

Panel 3. Dep. Variable: Takes Exit Exam

	(1)	(2)	(3)	(4)
RD_Estimate	0.006**	-0.003	-0.006	0.006
	(0.002)	(0.003)	(0.004)	(0.005)
Mean Just Below Cutoff	.019	.024	.013	.004
Optimal Bandwidth	66.9	47.2	47.0	34.3
Cutoff Value	500	600	700	720
Effective Observations	117,261	55,975	16,288	8,034
All Observations	250,758	250,758	250,758	250,758

Panel 4. Dep. Variable: Takes Teacher Evaluation

	(1)	(2)	(3)	(4)
RD_Estimate	0.004***	0.003*	-0.001	-0.005
	(0.001)	(0.001)	(0.001)	(0.003)
Mean Just Below Cutoff	.004	.006	.002	.005
Optimal Bandwidth	68.2	62.2	44.2	23
Cutoff Value	500	600	700	720
Effective Observations	119,378	73,348	15,287	5,279
All Observations	250,758	250,758	250,758	250,758

Panel 5. Dep. Variable: Exit Exam Score

	(1)	(2)	(3)	(4)
RD_Estimate	-0.042	0.076	-0.151	0.958
	(0.078)	(0.098)	(0.369)	(0.667)
Mean Just Below Cutoff	-.202	.415	1.157	.58
Optimal Bandwidth	63.6	58.9	17.6	16.7
Cutoff Value	500	600	700	720
Effective Observations	2,330	1,543	60	28
All Observations	4,319	4,319	4,319	4,319

Notes: Table 4 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variables are Graduation, Dropout, Employment at Schools and Taking the Exit Exam in Panels 1 through 4, respectively. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for high school GPA and the demographics described in Table 2.

Taken together these results indicate that the BVP policy increased the number of higher scoring students in teacher colleges, who went on to work in schools eight years later. This finding indicates that the policy was successful at raising the predicted quality of students who entered into the teaching profession.

The rest of early productivity indicators in Table 4 suggest that the higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores. This finding is most useful from a policy perspective, because it suggests that the predicted relationship between pre-college academic achievement and teacher medium run outcomes is invariant, and can be used in policy design.

Policy effects depend crucially on other policies

In this section we expand our results on teacher college enrollment for additional cohorts of test takers and also show results for the period when the BVP was not yet implemented.

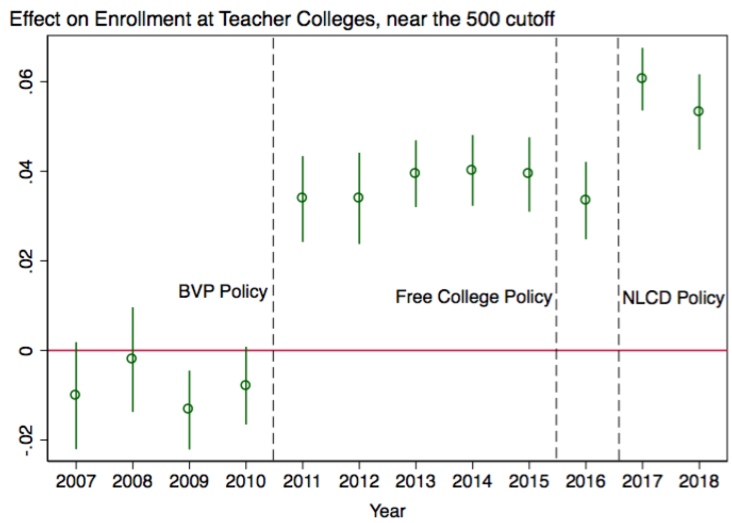
Figure 10 and Table 5 show results for the cohorts 2008 through 2018. We first focus on effects on the 500 cutoff, shown in Figure 10a. As a falsification test, we show that there are no effects before 2011. That year, when the BVP was implemented, enrollment at teacher colleges jumped 3.2pp. The effect, computed from RDs for every year, stays similar until 2017, where a new policy –called NLCD– was implemented. We discuss this policy and its results in detail in the next section.

Figure 10b shows how the effects near the 600 points threshold do depend on the existence of other financial 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 the free college policy was implemented. Figure 10b shows that the effect decreased considerably over time to zero in 2016.

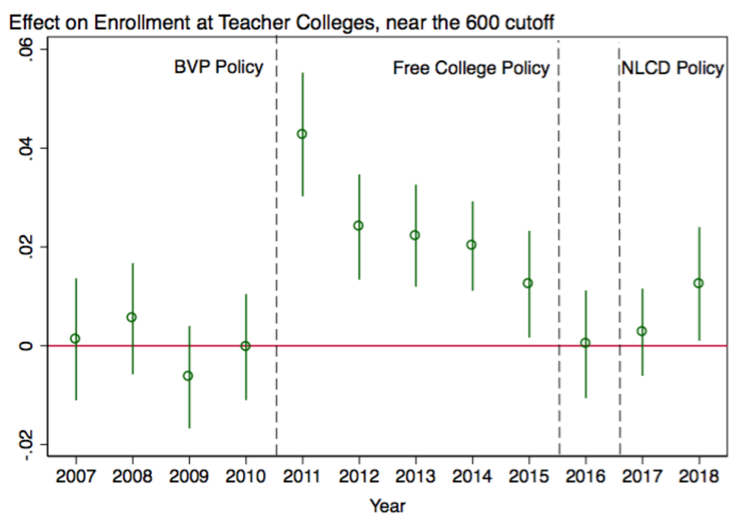
Consistent with our previous results we find no effects for higher scoring test takers, as we show in Figure 10c and Table 5. This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high achieving students have many good alternatives, so it is harder to move them towards teaching.

Figure 10: Effects on Enrollment over Time

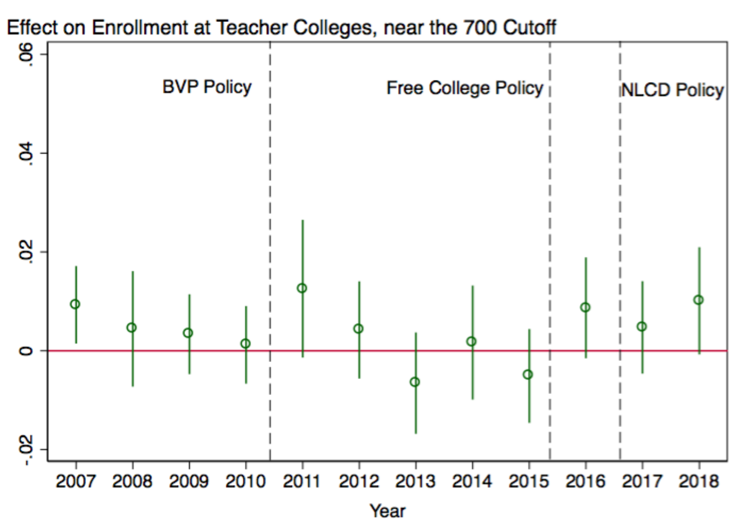
(a) Figure 10a



(b) Figure 10b



(c) Figure 10c



Note: Figure 10 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600 and 700 cutoffs, in Figure 10a, Figure 10b, Figure 10c and respectively. The dependent variable is Enrollment at Teacher Colleges for every regression. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for high school GPA and all the demographics described in Table 2.

Table 5: BVP Effects on Enrollment over Time 2008-2018

Panel 1. Dep. Variable: Enrollment at Teacher Colleges near the 500 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	-0.002	-0.013***	-0.008*	0.034***	0.034***	0.039***	0.040***	0.039***	0.033***	0.061***	0.053***
	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\widehat{\alpha_0}$	0.125***	0.130***	0.134***	0.085***	0.087***	0.068***	0.058***	0.052***	0.063***	0.042***	0.051***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Eff Size	-.016	-.102	-.059	.396	.39	.58	.695	.752	.532	1.451	1.044
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	500	500	500	500	500	500	500	500	500	500	500
N	77,865	87,108	90,169	90,450	84,773	86,341	86,955	90,065	90,725	93,455	97,357

Panel 2. Dep. Variable: Enrollment at Teacher Colleges near the 600 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.005	-0.006	-0.000	0.043***	0.024***	0.022***	0.020***	0.012**	0.000	0.003	0.013**
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.006)
$\widehat{\alpha_0}$	0.094***	0.098***	0.093***	0.096***	0.087***	0.074***	0.069***	0.074***	0.070***	0.068***	0.075***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Eff Size	.058	-.065	-.003	.448	.277	.3	.291	.169	.004	.04	.168
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	600	600	600	600	600	600	600	600	600	600	600
N	52,485	58,302	60,345	59,437	59,044	60,076	59,428	64,005	60,442	62,200	64,579

Panel 3. Dep. Variable: Enrollment at Teacher Colleges near the 700 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.004	0.003	0.001	0.013*	0.004	-0.007	0.002	-0.005	0.009*	0.005	0.010*
	(0.006)	(0.004)	(0.004)	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
$\widehat{\alpha_0}$	0.013***	0.012***	0.013***	0.030***	0.024***	0.031***	0.034***	0.027***	0.024***	0.023***	0.022***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Eff Size	.339	.272	.094	.423	.175	-.209	.049	-.19	.36	.208	.451
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	700	700	700	700	700	700	700	700	700	700	700
N	15,426	17,509	17,775	17,586	18,692	18,403	18,097	18,405	17,556	17,677	18,864

Panel 4. Dep. Variable: Enrollment at Teacher Colleges near the 720 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.002	-0.002	0.007*	-0.007	0.006	0.004	0.002	0.013**	-0.003	0.001	0.006
	(0.005)	(0.004)	(0.004)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
$\widehat{\alpha_0}$	0.008**	0.010***	0.006*	0.027***	0.022***	0.022***	0.026***	0.018***	0.023***	0.022***	0.020***
	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Eff Size	.235	-.232	1.327	-.256	.253	.183	.071	.759	-.114	.032	.289
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	720	720	720	720	720	720	720	720	720	720	720
N	10,720	12,166	12,630	12,488	13,123	12,864	12,496	12,755	12,270	12,275	13,275

Notes: Table 5 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variable is Enrollment at Teacher Colleges for every regression. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for high school GPA and all the demographics described in Table 2.

5 A Mandatory Screening Policy

This section studies a screening policy implemented in 2017, as part of the NLCD Law (*Nueva Ley de Carrera Docente*)[14]. 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 colleges must also comply with new admissions screening rules based on pre-college academic achievement.

In this section we explore the consequences of the screening policy. First, we compute regression discontinuity estimates near the college admission cutoff of 500 points. We can do this with the most recent data for enrollment in years 2017 and 2018. The difference at the discontinuity between across years informs of the NLCD effect, which interacts with other policies as we show in Figure 10.

In a next exercise we apply the screening rule to prior cohorts and determine what the partial equilibrium effects would have been in terms of total enrollment and later teacher outcomes. In this exercise we expect the screening 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.

In a third exercise we explore whether a more flexible data-driven rule would be better at screening future teachers. We evaluate how successful our data driven method can be –compared to the screening criteria proposed by the government– at minimizing the mistakes of rejecting future effective teachers.

5.1 NLCD Policy Specifics

The requirements for the screening policy affects admissions to all teacher colleges and are designed to be implemented gradually. During the first three years (2017-2019), the screening policy (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 entrance 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 entrance exam scores above 550 points (the 70th percentile) or be in the top 10% of their cohort GPA. If the student is in the top 30% of the GPA distribution at their high school and has scores above the average, then that student can also matriculate at teaching colleges.

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

[14]The NLCD is available here.

5.2 Effects on Teacher College Enrollment near the 500 points threshold

The first figure in Figure 10, Figure 10a, shows the effects generated at the 500 point cutoff. Given that the BVP policy was already implemented, the difference between the coefficients in 2016 and 2017 gives us the NLCD effect. Table 5 shows that the coefficient for 2016 was 3.3 pp and the same coefficient jumped to 6.1 in 2017. These parameters suggest that the NLCD policy decreased by half the fraction of low scorers who got enrolled at teacher colleges. The same table shows, as falsification tests, that no effects appeared for the other cutoffs of 600, 700, and 720 points.

5.3 Simulating the Screening Rule Back In Time

We simulated the policy rule of 2017 backwards to years 2007 to 2016. In Figure 11a we show the number of students that would have been accepted and the share of students who would have been rejected if the policy was implemented.

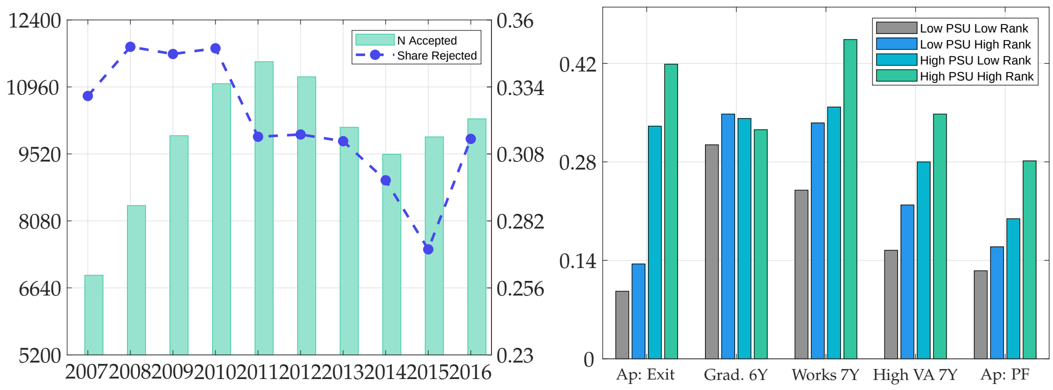
The number of accepted students would have continue decreasing as well as the proportion of students that would have been rejected by the screening rule. According to the Figure, the share of students who would have been rejected was between 25 and 31% in 2007 and remained at this range until 2010. From 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 11b, 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 11b, 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.

Figure 11: Screening Rule Back in Time

(a) Rule Applied To Past Teaching Students

(b) Simulation on performance outcomes



Note: Figure 11 shows trends and outcomes for students that would have been admited by P17 from 2007 to 2016. In Figure 11a 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 11a shows the labor outcomes for each group of students enrolled in pedagogy from 2007 - 2016.

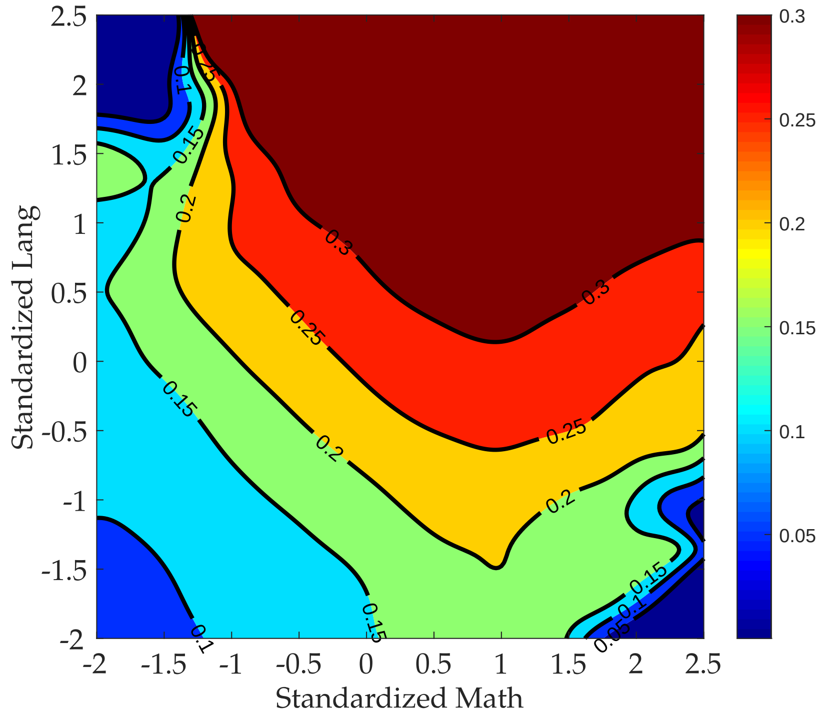
5.4 Towards A More Efficient Screening Policy

Increased availability of data and algorithms can help policy-makers make better predictions compared to previous section. Chile is a special case for exploring such data driven tools, because the MINEDUC has information systems that already produces information for the population of teachers in the country.

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 different subjects of the entrance exam and their GPA in their high school GPA.

Figure 12 highlights the potential role for a flexible screening rule. It shows the conditional mean off being hired by a school with high value added 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 12: Contour Plot of the Estimated Conditional Mean of Observational Teacher Evaluations

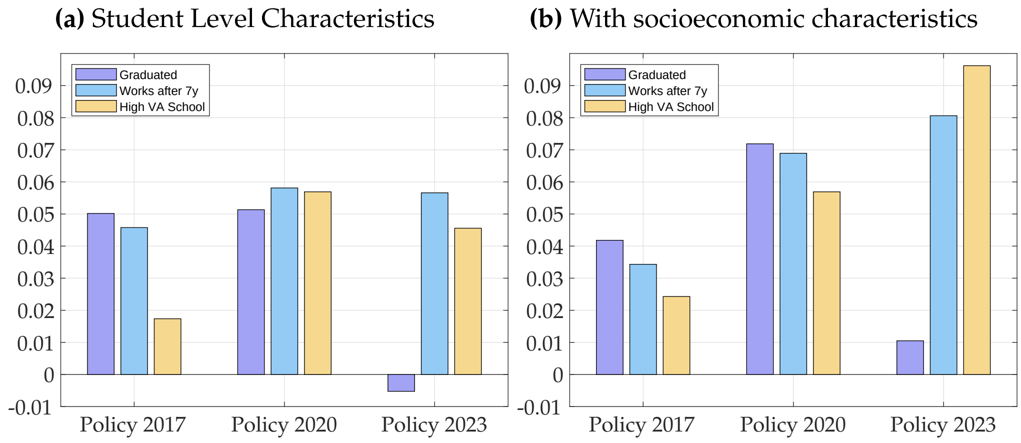


Note: This figure shows the contour plot of the mean of teacher evaluations conditional on pre-college academic achivement in math and language scores.

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 for students who enrolled at most 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 split in a training set (85%) and test set (15%) consisting of 24778 and 2754 observations respectively.

After cross-validating 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 cross-validation across different models, samples and outcomes are in the Online Appendix. Our AUC estimates are higher than the standard for predicting behavioral outcomes Chalfin et al. (2016).

Figure 13: Outcomes for Those Screened In Simulation ML



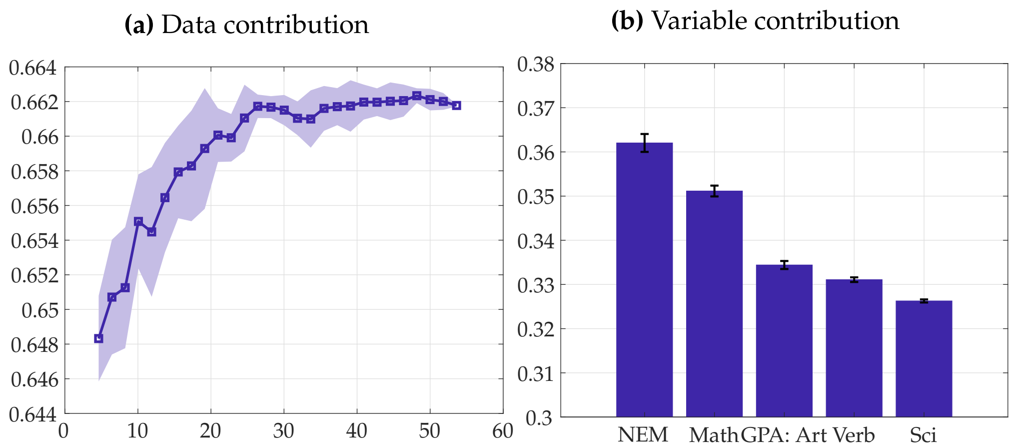
Note: The figures above show the percentage increase in each Graduation, working after 7 years and working in a good school for the students that would have been admitted by an ML screening method with a count of students rejected equivalente to those screened out by the rules proposed by the government. Figure 13b shows the results when using student level characteristics and school level characteristics such as socioeconomic status. Figure 13a 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 opposed 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 keeping the number of students admitted constant is to evaluate if the ML screening method has the potential to make less mistakes at (1) admitting prospective satisfactory teachers that would have been mistakenly rejected by (2) rejecting admission 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 13a shows that although the data driven method does not outperform P17 rule by much in the High Value Added outcome (an increase in performance of around 2%), it significantly improves selection of students graduating after 6 years and working after 7 years by 5 and 4% respectively. 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 6% and the number of teachers in schools with high value added in 5.5%. When compared to P23 which has a stricter rule, the improvement remains but only for working after 6 years and working in high value added schools in 6% and 4.5%. 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 13b 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 worse than the first method when compared to P17 but better when compared to P20 and P23 which are tighter rules and focus on high performing students. Adding socioeconomic variables increases the performance of the ML rule when compared to P20 in 2 percentage points the prediction of timely graduation and in 1 pp for students working after 7 years. The increase is higher, when compared against P23, adding 3 and 4 percentage points to our predictions of teachers working after 7 years and working in high value added schools respectively. Our results imply that socioeconomic variables can play an important role to predict success for teachers that are at the higher tail of the performance distribution.

Figure 14: Model accuracy composition



Note: Figure 14a 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 14b 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 14, we examine up to what extent is it possible to continue increasing the accuracy of our model. We plotted the area under the curve evaluated in the test sample for models trained with different levels of sample size in the training sample. Figure 14a 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 on performance of the model, Figure 14b 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.

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 with administrative records on the population of teachers in a middle income country 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 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 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, 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 policy increased the number of higher scoring students in teacher colleges, with the highest effects at the lower cutoffs of the college entrance distribution (about 37% of an effect size). This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high achieving students have many good alternatives, so it is harder to move them towards teaching.

Moreover, early productivity indicators measured eight 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 at the 600 cutoff 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. Our regression discontinuity estimates near this cutoff suggest that the policy screening out about half of the least academically prepared applicants near the threshold. 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 6% 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.

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, 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, Mara, "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 19972007," *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.

Appendix

Timeline of Higher Education Reforms in Chile[15]

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.

[15] 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