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

Christopher Neilson
Princeton University and NBER

Sebastian Gallegos
UAI Business School

Franco Calle
University of Chicago Booth

Mohit Karnani
MIT

February 4, 2022

This paper studies screening and recruiting policies that 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 use an RD design to evaluate two recent policies that increased the share of high-scoring students studying to become teachers. We then show how data-driven algorithms and administrative data can enhance similar teacher screening and recruiting policies.

*Neilson (corresponding author): Department of Economics, Princeton University, 20 Washington Road, Princeton, NJ 08544. Email: cneilson@princeton.edu. We thank seminar participants at the World Congress of the Econometric Society at Bocconi University, the NBER Education Group Fall Meeting, the Harvard Kennedy School, 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. We thank the Department of Evaluation, Measurement and Educational Registry (DEMRE) and Ministry of Education (MINEDUC) of the Government of Chile for access to the data and help locating and accessing historical data records. We thank Macarena Alvarado and Alvaro Carril for contributions to early versions of this paper. Gallegos also thanks the Center for Studies of Conflict and Social Cohesion (ANID/FONDAP/15130009) and dedicates this paper to the memory of the late Robert J. LaLonde. All errors are our own.

1 Introduction

Effective teachers matter for students' short and long run outcomes (Chetty et al., 2014; Araujo et al., 2016) and accordingly, governments aim to increase their teachers' productivity (OECD, 2005). A commonly used set of policies looks to increase the effectiveness of teachers once they are in the classrooms through incentives, training, accountability measures or rewards (see,.e.g. Biasi, 2021). 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).

Recruiting policies can be convenient compared to the on-the-job policies for several reasons. The first is that they 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 encourage effort (Hoxby, 1996; Hanushek, 2011; Biasi, 2021). [1]Third, the evidence suggests that later investments in training have a little influence on teacher productivity (Jackson, 2012; Lombardi, 2019).

The design of effective recruiting policies is hard. It requires a good prediction of teachers' effectiveness ex-ante, which has been elusive in the past (Harris and Sass, 2011; Jackson et al., 2014). Recruitment policies would ideally be informed by causal evidence, but prior research has been largely correlational in design (See et al., 2020). Administrative sources and historical records are being digitized and governments are developing the capacity to store and use the data (Figlio et al., 2017). This increasing data availability is likely to help overcome the lack of informative determinants of future teacher productivity and produce research designs helping to identify the causal effects of policies and initiatives on teacher recruitment. In addition, the development of improved predictive algorithms is lowering the cost of making more accurate predictions and influencing decisions, such as hiring, in many markets (Agrawal et al., 2018; Chalfin et al., 2016).

This paper studies policies that use pre-college achievement to recruit or screen out students entering teacher-colleges. We first show that teacher effectiveness might be predictable thanks to better data availability. We use recently digitized historical records from 1967 onward and link them to the population of teachers in Chile, to document the relationship between their own academic achievement at age 18 and their productivity as teachers up to 30-40 years later.

We then estimate the causal effects of two recent policies that both restricted and incentivized entry to teacher-colleges. Using administrative records and several regression discontinuities based on the policies' eligibility cutoffs, we assess whether they attracted higher-scoring test-takers to teacher colleges. We also examine later outcomes (like graduation, performance at exit exams and employment in schools) and how the policies interact over time.

Finally, we ask whether the combination of better administrative data and flexible prediction methods can enhance the recruitment procedures implemented by policymakers. We use machine learning methods to find that data-driven algorithms might outperform traditional cutoff-based mechanisms. This data driven approach seems promising for better targeting of investments in future teachers.

Our main findings are as follows. First, we show that there is a robust posi-

[1]Flexible pay schemes might even increase the gender wage gap (Biasi and Sarsons, 2021).

tive and concave relationship between teachers' pre-college academic achievement and a variety of short and long run teacher outcome measures. Our measures of teacher productivity include short run outcomes such as graduation from teacher colleges and college exit exams; and longer run outcomes such as wages, employment, external classroom teaching evaluations, students' achievement gains and students' perceptions about teaching effectiveness. 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 teachers' 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.

This evidence suggests that college entrance exams could be useful to screen out or recruit students entering teacher-colleges. We study two related policies implemented within the last ten years. The first policy started in 2011 and offered full tuition subsidies for test-takers with scores in the top 20% of the exam distribution. It also required participating teacher colleges to reject applicants with scores below the national mean. The second policy enacted in 2017 extended this requirement to all teacher colleges in the country.

We evaluate these policies using regression discontinuities based on the eligibility score cutoffs for high and low scoring applicants. We implement this empirical strategy using individual level data from the population of test-takers in the country, for ten cohorts of students.

Our findings show that the policies increased the number of higher scoring students enrolled in teacher colleges, with the largest effects at the lower cutoffs of the college entrance distribution (about 37% of an effect size). Effects at the higher cutoffs are large computed as effect sizes (about 100%) but small in levels. 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.

Eight years after the policy was first implemented, 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. 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 finally assess whether the use of data-driven algorithms may enhance the screening procedures planned for the future by the government. We train classification trees that outperform all the different government recruiting policies, by going beyond single-dimensional cutoff rules. Importantly, our classifiers are simple enough to not sacrifice interpretability nor relying on complicated sets of input features. Taken together, the findings support the use of machine learning methods as a promising way aiding screening and recruitment policies.

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 and coauthors (see,.e.g., Tincani, 2014; Tincani et al., 2016; Tincani, 2021) which explicitly models the sorting process and simulates related teacher policies. This paper complements this structural work with empirical descriptive and causal evidence of the relationship between pre-college academic achievement and later outcomes.

Our 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. 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 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.

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 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.

Finally, our findings highlight avenues for further research in an increasingly data-rich environment where prediction is a key input to policy design (see, e.g., Mullainathan and Spiess, 2017; Kleinberg et al., 2017). Newer methods are being implemented to exploit increasing amounts of data, and we believe that empirical exercises similar to ours (e.g. Athey and Imbens, 2019; Athey, 2019; Sajjadiani et al., 2019) will be increasingly common in the near future.

2 Context, Policy and Data

2.1 Context

Chile is a middle income country that has reached low levels of teacher absenteeism and a 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 Latin America; Chaudhury et al. (2006) estimate absenteeism rates of 15% in Brazil, 14% in Ecuador, and 11% in Peru.

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 number of classroom teachers[2] has increased from 125,000 in 2008 to 164,000 in 2018 (MINEDUC, 2019), while student enrollment 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).[3]

With enough teachers in the classrooms and high rates of student enrollment (OECD, 2009), the policy focus in the last ten years has been devoted 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; Hanushek et al., 2019).[4] Consistently, we know from the related literature that college graduates with higher college entrance scores are less likely to enter teaching (Vegas et al., 2001; Hanushek et al., 2019; Estrada and Lombardi, 2020), and Chile is no exception.

Figure 1 shows that in year 2010 (before the implementation of teacher recruiting policies described below), teacher colleges' students scored only 0.1 standard deviation (σ) above the national mean in the college entrance exam,[5] while students enrolled in other fields like engineering, law, and medicine scored about 0.6σ above. Also, test scores for education students had been declining over time, since in 1995 students from teacher colleges scored 0.3σ over the national mean (Alvarado et al., 2011). 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).

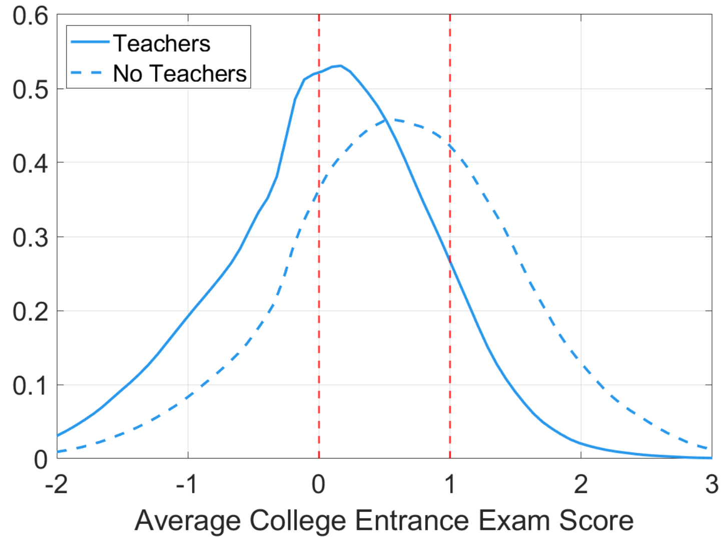
[2]Teachers in Chile work in public schools, which are funded and administered by the government; voucher schools, that are funded mainly with public funds but administered by privates; and private schools are both funded and administered privately.

[3]These numbers are consistent with the demographic transition being experienced by Chile, exhibiting low fertility and mortality rates, and relatively high life expectancy (World Bank, 2011).

[4]Mizala and Nopo (2016) estimate a teacher underpayment of about 20% in Latin American countries in 2007 (with a 18% for Chile) after controlling for a set of characteristics linked to productivity. Hanushek et al. (2019) estimate that teachers in the U.S are paid some 20 percent less than comparable college graduates. Evans et al. (2020) find that, in 7 out of 15 African countries, teachers suffer a deficit in earnings relative to comparable wage workers that averages 26%.

[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



Note: Figure 1 plots the distribution of college entrance exam scores before the implementation of teacher recruiting policies, for two groups: freshmen in teacher colleges (continuous line) and freshmen in the health, law and STEM fields (dotted line). The entrance exam score (in standard deviation units) is the average of the math and language exams. We provide further details of the college entrance exam in section 2.3.

2.2 Teacher Recruitment Policies

In this context, the Chilean government implemented two policies to recruit talent at teacher colleges. The first policy, called *Beca Vocacion Profesor (BVP)* was implemented in 2011. It consisted in full tuition subsidies for prospective students who scored about 1σ above the mean in the college entrance exam. The BVP policy also required participating teacher colleges to reject applicants with scores below the national mean.

The second policy was the *Nueva Ley de Carrera Docente* (NLCD) and started in 2017. The NLCD basically imposed the BVP restrictive requirement for admissions at all teacher colleges across the board. Under this policy all applicants to teacher colleges had to have college entrance exam scores at least as high as the national mean, or have a high-school GPA in the top 30% of their graduating cohort. We provide more specific details and assess both policies in section 4.

2.3 Data

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. The historic records come from digital copies of old books and newspapers collected as a part of the work done in Hastings et al. (2014). In continued partnership with the national agency in charge (DEMRE) we complemented these data digitizing additional test scores back to the first test in 1967. For the more recent cohorts of test takers (2004-), the DEMRE has made electronic records available.

The Chilean national college entrance exam is similar to the SAT in the United 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 approximately 500 and standard deviation of 110. The exam scores are required to apply to all public universities and most private universities and institutes.

Data on Teacher Productivity. We gathered a host of teacher productivity proxies from different sources of administrative records. Our measures of teacher productivity include short run outcomes such as graduation from teacher colleges and college exit exams; and longer run outcomes such as wages, employment, external classroom teaching evaluations, students' achievement gains and students' perceptions about teaching effectiveness.

In the next section, we correlate all these measures of teacher productivity with the digitized pre-college achievement described above. While provide specific details in our *online appendix*, we describe each measure and data below.

Graduation from Teacher Colleges. We use microdata on the population of teacher college enrollment and graduation, which the Chilean Ministry of Education (MINEDUC) started to collect in 2004 and 2009, respectively. We constructed graduation rates for 105K individuals combining enrollment records from years 2004 to 2010 with graduation data for years 2009 to 2018. This procedure allows us to study graduation rates that were 'on time' (i.e., within 5 years after initial enrollment, at approximately 23 years old) and also late graduation (i.e., up to 8 years after enrollment, at about 26 years old).

Exit Exams. The exit exams were first implemented in 2009. Our data consists in microdata for all the exit exam's test-takers between 2009 and 2017. The sample consist of about 35K just graduated teachers with scores on different exams, like 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 on average about 25 years old.

Government Evaluations. The government started to implement teacher evaluations in public schools since 2004. We gathered information for 63K classroom teachers, evaluated between 2004 to 2017. Each evaluated teacher receives an overall score at the end of the evaluation process. The MINEDUC uses that score to classify teachers into four categories of performance, from best to worse: outstanding, competent, basic, and unsatisfactory. The overall score composed by four components: (i) a self-evaluation questionnaire (10%); (ii) a third-party reference report, filled by the school principal or supervisor (10%); (iii) one peer review (20%), and a teacher performance portfolio (60%) that collects direct evidence on teaching skills, pedagogical decisions and classroom practice.[7] Previous research (OECD, 2013; Bruns and Luque, 2015) suggests that the portfolio component has the strongest association with students' progress measured using standardized test scores. Therefore, we use both the overall score and the portfolio score in our analysis later on. On average, evaluated teachers were 40 years old at the time of the assessment.

[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] As we expand in the online appendix, the portfolio component includes two modules. In the first module, teachers plan a class defining its contents and related assessments. They are also asked questions about teaching practices. The second module consists in a videotaped class followed by a questionnaire on the students behavior and understanding, and the teacher's own performance.

Employment in Schools. We gathered information for about 240K graduates from teacher colleges in years 1995 to 2017 which we merged with the population of teachers working in schools between 2003 to 2018. We compute whether graduates worked during that period of time 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. The MINEDUC collected information on teacher wages by asking principals about teachers wages and working hours in year 2011. Teachers with information on wages are about 117K. Teachers working in public schools are approximately 40 percent of the sample (49K). They benefit from a special labor code, which makes wages grow with tenure and not expected to change with productivity. Teachers in voucher schools represent 60 percent of the sample (68K). 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. On average, teachers with wage information are 37 years old.

Students' Achievement Gain. We use students' achievement gain during the academic year as another proxy of teacher productivity. The MINEDUC does not implement value-added exams, but De Gregorio and Neilson (2020) implemented math tests especially designed to measure the gain in achievement for students with the same teacher during the academic year. The sample consists of about four thousand students in grades 9th to 11th tested at the beginning and the end of the year in 2016.

Students' Perceptions. We use students' perceptions regarding effective teaching as an additional measure of teacher quality. The survey implemented by De Gregorio and Neilson (2020) to the same four thousand students follow the recommendations of the Measures of Effective Teaching study carried out in the U.S. (Kane and Cantrell, 2010). Questions are categorized eight dimensions of teaching practices and classroom environment: positive culture and learning environment, student understanding checked for and ensured, engaging learning environment, expectations held by teacher, student input and ideas valued, learning fully internalized by students, encouraging and supporting relationships fostered, and classroom participation.

3 Pre-college Achievement and Teacher Productivity

In this section we document the 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.

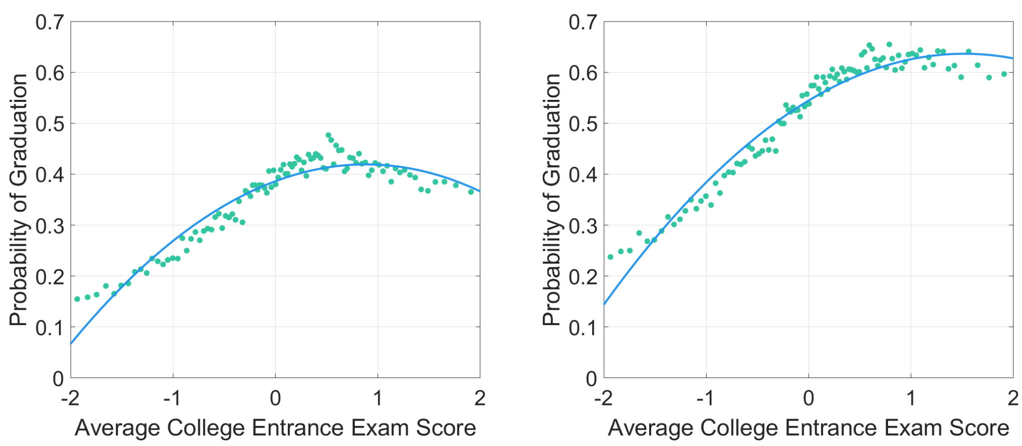
The general takeaway is that the empirical relationship between pre-college skills and own teacher productivity later on is positive and concave. We report regression coefficients in Table 1, organized in five panels of teacher outcomes: graduation, exit exams, productivity measures, labor market results and their students' outcomes. The coefficients come from separate regressions of different measures of teacher outcomes on the college entrance exam score (in standard deviation units and labeled 'PSU Score') and its square. The estimates on scores are all positive and significant, and most coefficients on the squared term are negative.

We discuss our estimation results complemented with simple visual evidence below. All figures in this section plot the y-axis variable within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data.

In Figures 2 and 3 we examine early outcomes of students from teacher colleges, like graduation rates and exit exams. Figure 2 shows that graduation rates correlate positively with test scores at entry, and that the relationship is concave. The results in graphs 2a, 2b and the first panel of Table 1 show than an increase in one standard deviation on the college entrance exam scores relates to an increase in graduation rates of 7.0 and 11.2 percentage points after 5 and 8 years of initial enrollment. These are increases of approximately 20% relative to the baseline graduation rates of 35% and 50%, respectively. The positive correlation is much flatter for scores above the mean.

Figure 2: College Entrance Exam and Graduation from Teacher Colleges

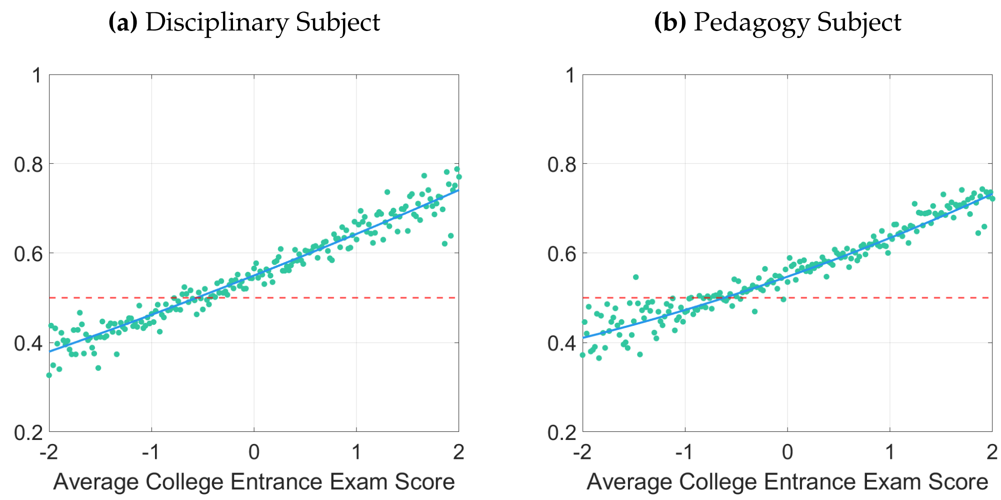
(a) Graduation within 5 years **(b)** Graduation within 8 years



Note: The figures plot the probability of graduation after 5 years (Figure 2a) and 8 years (Figure 2b) 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 2010 who graduated between 2009 and 2019. In both Figures the sample size is of $N = 105,422$.

Figure 3 shows the correlation between test scores at college entry and exit exams taken just before graduation. The graphs and the corresponding coefficients in Table 1 show that one standard deviation on the college exam test scores is associated to an increase of 0.50σ on the disciplinary and pedagogical skills measured in the exit exam. Table 1 also reports that one standard deviation in test scores is related to an increase of 0.28σ and 0.54σ in the writing skills and ICT skills exit exams, respectively.

Figure 3: College Entrance Exam and Teacher College Exit Exams

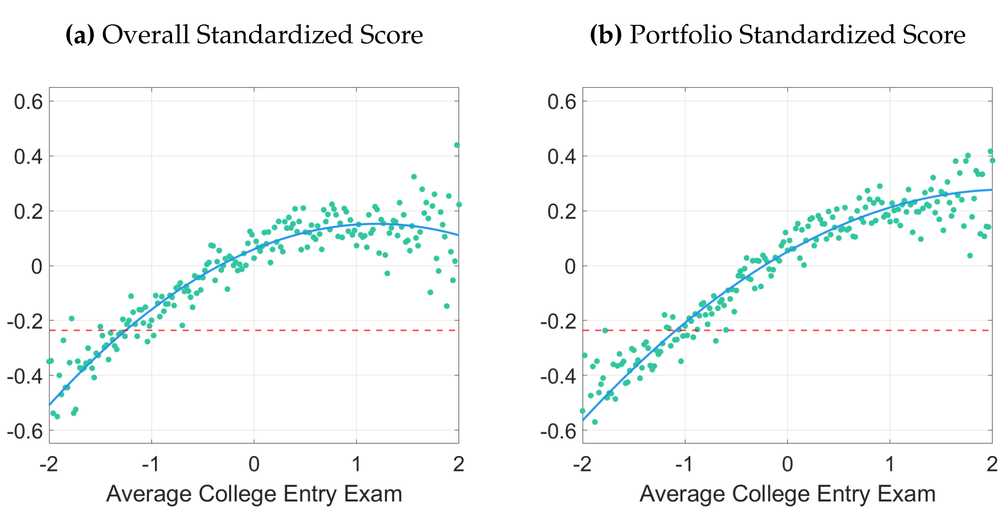


Note: The figures plot the standardized scores of two exit exams within 100 equal-sized bins of the average college entrance exam score (in std. dev. units) and fits estimated lines using all the underlying data. The two exams are disciplinary skills (Figure 3a) and pedagogical skills (Figure 3b). The data consists of exit exam test takers between years 2009 and 2017. The sample sizes are $N = 35,355$ in Figure 3a, and $N = 33,409$ in Figure 3b.

We now present results for later outcomes, when individuals are teaching and working in schools. Figure 4 describes the bivariate relation between college entry exams scores and teacher evaluations taken up to 30 years later. As in Figure 2, the relationship is concave, 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 is linked to increases of 0.14σ and 0.19σ on the overall and portfolio evaluations scores, respectively.

Consistent with the concave relationship for the scores, Table 1 shows that one standard deviation in the entry exam scores is associated to a drop of 18% (5 percentage points over a mean of 28 percent) in the the likelihood of being classified as basic or unsatisfactory. Similarly, one standard deviation is related to a relatively smaller increase of 7% (5 percentage points over a mean of 72 percent) in the probability of being outstanding or competent.

Figure 4: College Entrance Exam and In-Class Teacher Evaluation



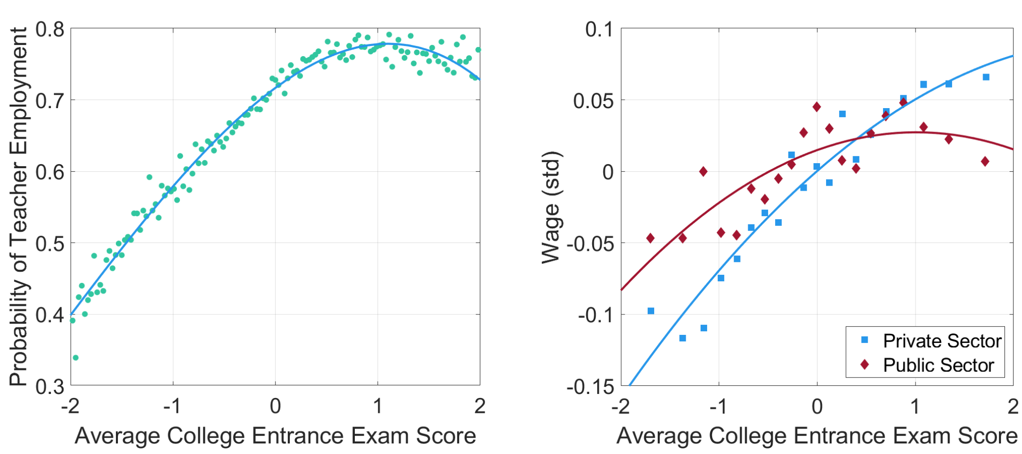
Note: The figures plot the teacher evaluation scores (overall in Figure 4a and the portfolio component in Figure 4b), within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in teachers evaluated between years 2004 and 2017. In both Figures the sample size is of $N = 63,539$.

Figure 5 exhibits correlations between the entrance exam scores and labor market outcomes, which are consistent with the concave productivity patterns described so far. Figure 5a plots the probability of working in schools for graduates from teacher colleges versus their college entrance scores. An increase of one standard deviation in scores increases the likelihood of working as a teacher in 10 percentage points (pp.) relative to a baseline of 68%, accompanied by a negative coefficient of 4 pp in the square of the scores. This result suggest that a fraction of teachers in the right tail of the distribution of college preparedness quit the profession by that time.

Figure 5b shows how hourly wages vary with scores, for 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 occur in the base of seniority. Consistently, the coefficients in Table 1 show that a standard deviation increase in scores is associated to 0.064σ and 0.024σ of hourly wages for teachers working in the private and public sector respectively (p-value of the difference=0.0001). The magnitude of the coefficient over wages is nearly 3 times higher for the sample of teachers in the private sector, where schools can adjust salaries almost unrestrictedly as teacher productivity changes. The same dynamic does not occur in the public sector where wages are much less flexible and determined primarily by years of service (seniority).

Figure 5: College Entrance Exam and Labor Market Outcomes

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



Note: Figure 5a and Figure 5b plot the fraction of teachers employed and their wages (in standard deviation units), respectively, within equal-sized bins of the average college entrance exam score, and fits estimated lines using all the underlying data. The data in Figure 5a consists in $N = 240,549$ graduates from teacher colleges in years 1995 to 2017, who are employed (or not) between 2003 to 2018. The data in Figure 5b consists of $N = 117,105$ teachers working in private and public schools in 2011.

The bottom panel of Table 1 shows that students' outcomes (like math gains and perceptions about teaching effectiveness) are also positively related to their teachers' entrance exams. The first three columns in the panel show that an increase of one standard deviation in the teacher PSU score is associated with an increase of 0.43σ to 0.29σ in gains in algebra, numbers and geometry. Results are also suggestive of a concave relationship though the negative coefficients are not precisely estimated.

The results for students' perceptions, in the last column of the bottom panel, follow the same pattern. We use a factor analysis to produce an index for student perception using the eight categories of teacher effectiveness reported by students.[8] One standard deviation in PSU scores is associated to 0.08σ in the students' perceptions index, with a negative and significant coefficient on the square of the PSU score. These results suggest that the relation between students' outcomes and teacher entrance exams are positive and concave, consistent with the previous set of teacher outcomes.

Table 1: College Entrance Exam and Teacher Outcomes

	(1)	(2)	(3)	(4)
Graduation	Years After Enrollment 5 Years	8 Years		
PSU Score	0.070***	0.112***		
	(0.001)	(0.001)		
(PSU Score)^2	-0.029***	-0.029***		
	(0.001)	(0.001)		
Observations	105,422	105,422		
Dep. Var. Mean	0.350	0.498		
Exit Exams	Disciplinary Test	Pedagogical Test	Writing Test	Technology Test
PSU Score	0.493***	0.505***	0.282***	0.539***
	(0.005)	(0.005)	(0.009)	(0.014)
(PSU Score)^2	0.043***	0.033***	-0.019***	-0.043***
	(0.003)	(0.003)	(0.006)	(0.011)
Observations	35,355	33,409	11,300	5,517
Dep. Var. Mean	0.000	0.000	0.000	0.000
Teacher Evaluation	Overall Score	Portfolio Score	Basic or Unsatisfactory	Outstanding or Competent
PSU Score	0.143***	0.189***	-0.051***	0.051***
	(0.004)	(0.004)	(0.002)	(0.002)
(PSU Score)^2	-0.050***	-0.037***	0.020***	-0.020***
	(0.003)	(0.003)	(0.001)	(0.001)
Observations	63,539	63,539	63,539	63,539
Dep. Var. Mean	0.000	0.000	0.283	0.717
Labor Market	Employment	Wages	Private Wages	Public Wages
PSU Score	0.096***	0.047***	0.064***	0.024***
	(0.001)	(0.003)	(0.004)	(0.005)
(PSU Score)^2	-0.036***	-0.003	-0.002	-0.010**
	(0.001)	(0.002)	(0.003)	(0.004)
Observations	240,549	117,105	67,909	49,196
Dep. Var. Mean	0.679	0.000	-0.000	0.000
Student Outcomes	Δ Algebra Tests	Δ Numbers Tests	Δ Geometry Tests	Perceptions Index
PSU Score	0.377***	0.292***	0.433***	0.077***
	(0.060)	(0.075)	(0.072)	(0.016)
(PSU Score)^2	-0.053	-0.079	-0.078	-0.031**
	(0.048)	(0.061)	(0.053)	(0.012)
Observations	3,756	3,756	3,756	3,612
Dep. Var. Mean	0.000	0.000	0.000	0.000

Note: Table 1 reports results from 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 five panels: graduation, exit exams, productivity measures, labor market outcomes and student outcomes. All results in panels 1-4 come from estimations at the teacher level and include year and teacher specialization fixed effects. Results in panel 5 come from estimations at the student level, with standard errors clustered at the classroom level and controls for teacher experience, class size, school socioeconomic status, and type (public or private). Robust standard errors are in parentheses. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively.

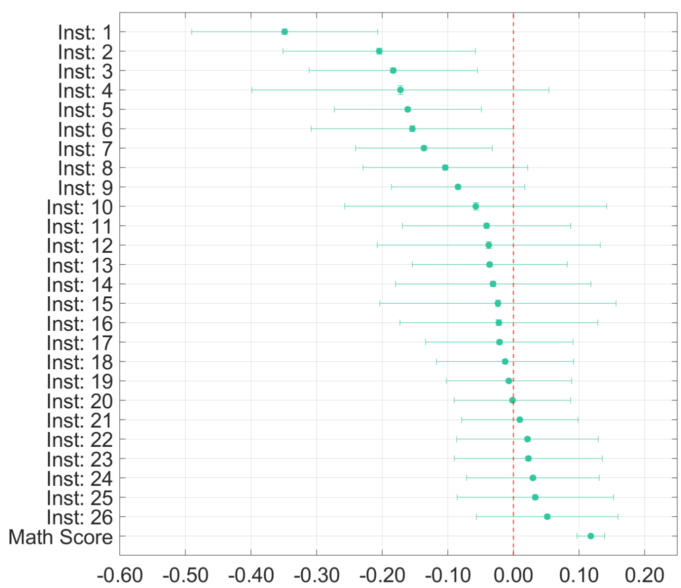
[8] In our online appendix, we report the results for each category individually, all of which show results that are very similar to the estimate for the index.

3.1 Teacher Colleges' Value Added

We end this section with a complementary exercise examining whether access to higher value-added teacher colleges causes these observed scores-productivity 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. While we document the exercise in detail in our *online appendix*, we show our main findings in Figure 6.

Figure 6 plots the regression discontinuity estimates for each institution on (overall) teaching evaluation scores as a proxy for value added. The results indicate that the institutions' value added cannot be distinguished from zero for most teaching colleges. This result suggest that teaching colleges are not adding differential value to the predicted gap in teacher effectiveness.

Figure 6: Education Institutions Value Added to Teacher Evaluation



Note: Figure 6 plots the regression discontinuity estimates on teacher evaluation scores by institution, using data on the population of applicants to teaching colleges from 1977 to 2011. We provide further details in our online appendix.

Taken together, the findings of this section indicate that below average pre-college achievement is systematically associated with lower performance as teachers measured up to thirty and forty years later. Colleges do not seem to be generating the correlation and therefore there might be space to use it for policy.

In the next section we study the causal effect of recruiting policies that used pre-college scores to attract higher achieving students to teachers colleges.

4 Assessing Teacher Screening Policies

4.1 A *Carrot & Sticks* Approach to Recruiting and Screening

The *Beca Vocacion Profesor* (BVP) used college entrance exams to recruit and screen out students entering teacher-colleges. We assess the policy on teacher college's enrollment and medium run outcomes like graduation, exit exams and employment in schools, all measured up to eight years after first implemented.

The 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 suggests that the policy was successful at raising the predicted quality of students who entered into the teaching profession.

4.1.1 BVP Policy Specifics

The BVP policy started in 2011 and offered full scholarships, stipends and paid semesters abroad for high scoring test-takers who enroll as freshmen at teacher colleges.[9]

Test-takers with scores in the top 20% (i.e., 600 points or more) were eligible for a full tuition scholarship.[10] Those with scores at approximately the top 5% (700 points or above) were eligible for the full tuition scholarship plus a monthly stipend of about $US150, which was close to 50% of the minimum wage. The top 2% scorers (720 points or higher) would benefit from the tuition, stipend and a paid semester abroad at a prestigious teaching college. For instance, advertisements mentioned a semester abroad at Stanford or in Finland.

The policy also 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.[11] 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).

4.1.2 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 *online 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.

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 the test-taker i. Our parameter of interest is α_1, which is an *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 scored above a particular threshold and zero otherwise. For simplicity, we estimate separate regressions for the 500, 600, 700 and 720 policy

[9] In practice, the only requirement to be eligible was taking the entrance exam on December 2010 aiming to start as a new first year student at a teacher college in March 2011. Students already enrolled in teaching careers were not eligible for the scholarship.

[10] 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.

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

cutoffs.[12] $f(S_i)$ 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 whether they attended a public or private high-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 population of test-takers in the country. 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.[13] A total of 250,758 high school graduates[14] took the college entrance exam in December of 2010, aiming to start classes at the beginning of the academic in March of 2011. All of these test-takers were potentially eligible for the BVP had they achieved scores above the policy cutoffs. The scores on the each subject (mathematics, language, history and science) have a mean of about 500 points. The college entrance exam score is the math-language average score. The math and the language tests are mandatory for all test-takers, while the history and science tests are optional exams.

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 these statistics are consistent with data coming from national censuses and surveys (CASEN 2016). About 55%, 35% and 10% of the test takers 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 in Table 2 shows the fraction of test takers who enroll in higher education after the exam (in March 2011). A 63% enrolls at any institution, 44% enrolls at colleges and half of that enrolls at the more selective universities.[15] About 20K test takers (8% of the total) enroll at any teacher college and approximately 8K (a 3% of the total) enroll at teacher colleges that were BVP-eligible.

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

[13] 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.

[14] The college entrance exam take-up among high school graduates is high. Each year, about 260K students graduate from high-school in Chile. Test-takers are typically a mix of just graduated high-schoolers (75%) and graduates from previous years (25%).

[15] These selective universities are non-profit institutions, grouped in the Council of Rectors of the Universities of Chile (CRUCH), which receive students with highest scores 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 at Test (years)	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 Selective College	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 math and the language tests are mandatory for all test-takers, while 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 variable 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. The Capital City variable 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, indicating whether individuals were enrolled during the academic year 2011. 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 selective does the same if the test taker enrolled at universities belonging to the *Consejo de Rectores*, a group of non-profit institutions that enroll the students with highest scores in the country. 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.

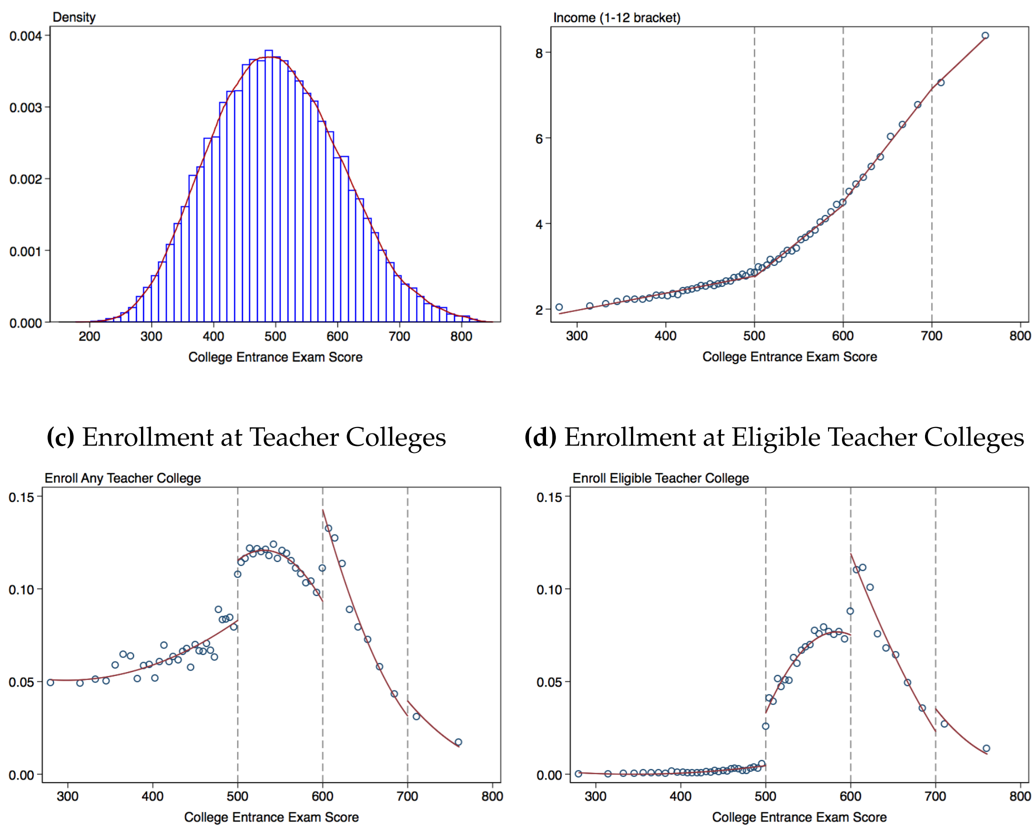
4.1.3 Results

Our main results show that the policy attracted higher scoring test-takers to teacher colleges. Figure 7 summarizes the first set of findings. Figure 7a and Figure 7b 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 7c and Figure 7d 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 different likelihood of enrolling at teacher colleges.

Figure 7: Main Results

(a) Score Distribution **(b)** Household Income

(c) Enrollment at Teacher Colleges **(d)** Enrollment at Eligible Teacher Colleges



Note: Figure 7a plots the distribution of scores for all test takers. Figure 7b, Figure 7c and Figure 7d 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 7 is of N=250,758 observations.

Table 3 provides the regression analog of graphs 7c and 7d in panels 1 and 2. The columns report the RD estimates from Equation 1 at the 500, 600, 700 and 720 cutoffs, with MSE-optimal bandwidths (Cattaneo et al., 2018) for each threshold. These are our preferred estimates, which are robust to different bandwidths and specifications, as we show in the *online appendix*.

The estimates from Panel 1 show sizable effects near cutoffs. The magnitude of the estimates represent relative increases of 37% at 500 points (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). We find a precise null effect at the highest cutoff of 720 points.

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.

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. All regressions control for the demographics described in Table 2. These are our preferred estimates, which are robust to different bandwidths and specifications, as we show in the *online appendix*.

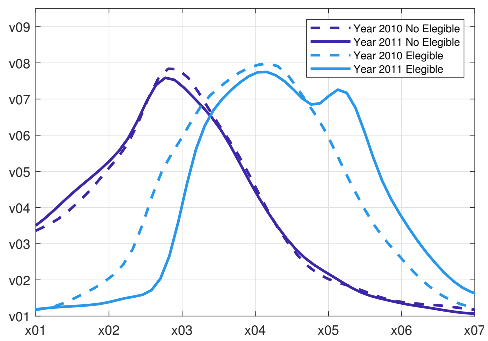
4.1.4 Approximating Aggregated Effects

We now add additional microdata data for the population of test-takers in 2010 to approximate impacts beyond the local effects estimated above. We use the policy time variation to compare outcomes along the distribution of test scores before and after the BVP was implemented.

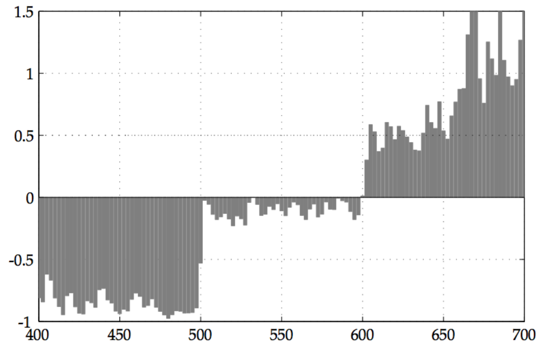
Figure 8 motivates the analysis with visual evidence showing how the policy shifted the distribution of scores in teaching colleges. Figure 8a shows distributions of entrance exam scores for the 2010 cohort (before the BVP) and the 2011 cohort (after), by enrollment at eligible and non-eligible teacher colleges. The distribution of scores at eligible teaching colleges shifts markedly to the right after the policy, while it remains the same for students at non-eligible institutions.

Figure 8: Aggregate Effects on the Distribution of Scores

(a) Scores at eligible and non-eligible teacher colleges, before & after the BVP



(b) Change in Probability of Choosing a Teacher College at Eligible Institutions



Note: In Figure 8a the continuous (dotted) line shows the distribution before (after) the BVP policy. The ■ color depicts the distribution for non-eligible colleges while the ■ does the same for eligible colleges. Figure 8b shows the before-after change in the probability of enrolling in an eligible teacher college conditional on enrollment, along the test score distribution.

Because all the action takes place at eligible institutions, we also examine the before-after change in the probability of choosing teaching, conditional on enrollment. We plot the results along the distribution in Figure 8b. The figure illustrates an increase close to 40% in the probability of enrollment at an eligible teaching college around 600 points, which increases above 100% at 700 points. Under 500 points the probability decreases by almost 100%, because students with scores below that threshold could not enroll using the BVP policy.

In the *online appendix*, we complement this analysis with a simple difference in differences model using data for groups of test-takers with less than 500 points, 500 to 600 points and 600 and more.

Our findings show that the BVP policy was successful in screening out lower scoring test-takers and attracted higher scoring applicants. However, is is useful to put the results in perspective. While the results show that the BVP raised choice probabilities significantly, the number of students at those margins is still relatively modest compared to the population of students in teaching colleges. We estimate that about 1,000 additional students entered teaching colleges from the top 30% of the distribution and that the screening restrictions reduced the bottom tail of the distribution by about 4,000. While meaningful, these figures must be considered taking into account that the total number of freshmen students at teachers' colleges is close to 20,000.

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

Table 4 reports our estimates. The estimates in Panel 1 show that the policy increased employment at schools of the higher scoring test-takers near the cutoffs of 500 and 600 points. The effect sizes are of 12% at 500 (1.2pp over 6.4pp) and 34% at 600 (2.3pp over 6.7pp).

Panels 2 to 4 present estimates on college 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 null effects suggest that higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores.[16]

[16]The take-up rates of the exit exam and the teacher evaluation are very low because these exams where not mandatory for the cohort of test takers under analysis. In the *online appendix* we show that there are no effects the exit exam score but estimates are much noisier due to the low number of observations.

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

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 Employment at Schools, Graduation, and Taking the Exit Exam and Teacher Evaluation 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 all 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.

4.2 Policy Effects Over Time

In the previous section we focused our analysis on the first cohort that benefited from the BVP policy. In this section we study the BVP policy effects for different cohorts over time, and examine how the results change when new policies are introduced.

We report results for ten different cohorts, from 2008 to 2018 in Figure 9 and Table 5. For each cohort, we estimate equation Equation 1 and report regression discontinuity estimates near the 500, 600 and 700 cutoffs.

Our main findings indicate that the effects of setting the minimum scores at the 500 cutoff remain high and persistent for cohorts over time. Second, that effects at the 600 cutoff tend to vanish when another 'free college' policy kicks in; and third, that there is essentially no action at the top (at the 700 cutoff) no matter what policy was in place.

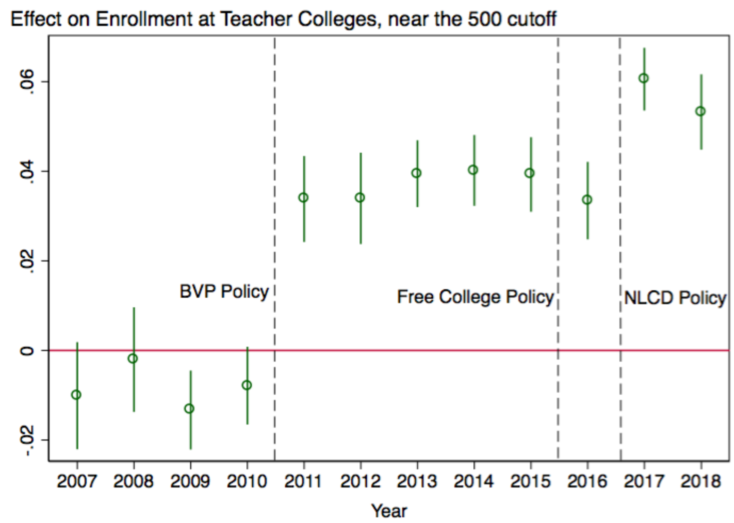
We plot the effects near the 500 cutoff over time in Figure 9a. As expected, the figure shows no effects for cohorts 2008 to 2010, before the BVP policy was implemented. Once the BVP was implemented, in 2011, enrollment at teacher colleges jumped 3.2pp, as we described in the previous section. The magnitude of this effect is similar for the next five cohorts until 2017, when the NLCD policy (described in section 2) was first implemented. We discuss NLCD policy and its results in detail in the next section.

Figure 9b plots the RD estimates near the 600 points threshold. As in Figure 9a, the figure shows zero effects before the BVP policy was implemented and positive effects after. In this case, effects diminish for the more recent years and disappear in 2016. The country started with a nation-wide policy to make tuition free, that was fully implemented in 2016. This *free college* policy appears to naturally have reduced the financial incentives generated by the BVP. Consistently the regression discontinuity estimates show that the effectiveness of the policy was significantly diminished for the newer cohorts. These results are aligned with contemporaneous work by Castro-Zarzur et al. (2019) and Castro-Zarzur and Mendez (2019).

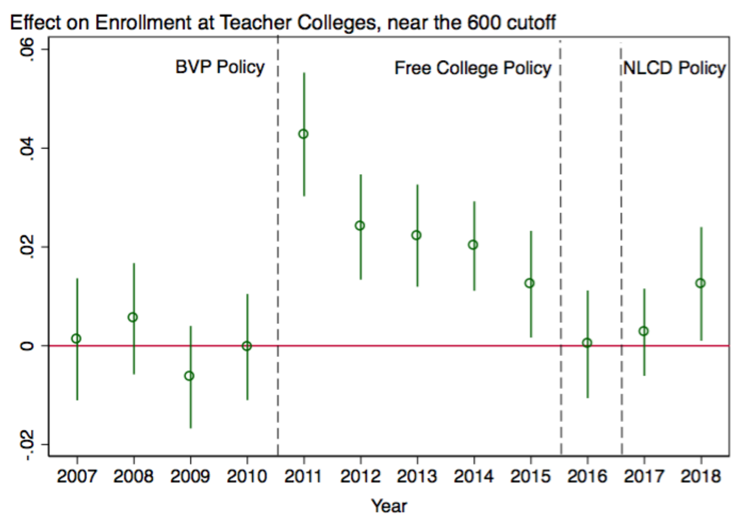
Finally, we find precisely estimated zero effects for higher scoring test takers near the 700 and 720 cutoffs. Figure 9c graphs the estimates for the 700 threshold, while Table 5 presents the estimates for both 700 and 720 cutoffs. This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high achieving students have many good alternatives, and it is harder to move them towards teaching careers.

Figure 9: Effects on Enrollment over Time

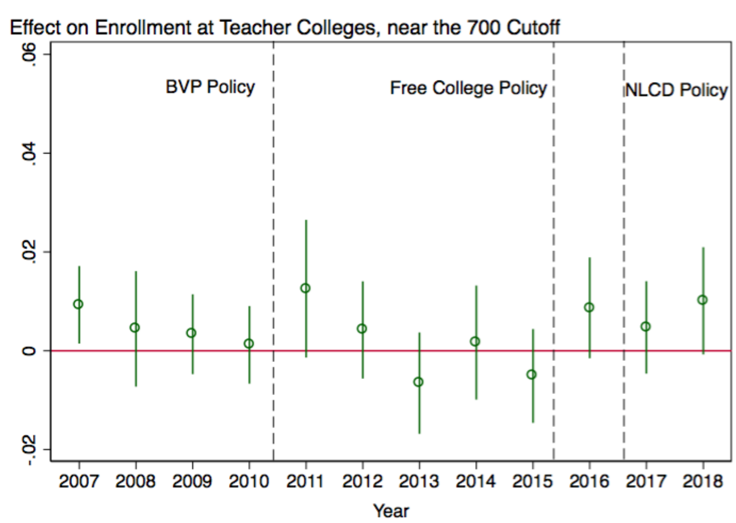
(a) Figure 9a



(b) Figure 9b



(c) Figure 9c



Note: Figure 9 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600 and 700 cutoffs, in Figure 9a, Figure 9b, Figure 9c 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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	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)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	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.

4.3 A Mandatory Screening Policy

The NLCD (*Nueva Ley de Carrera Docente*)[17] was enacted in 2017. While the screening component of the BVP policy prevented *participating* teacher colleges to admit applicants with scores below the national mean, the NLCD policy extended the

[17]The NLCD is a broad policy aimed to enhance the system of professional development for teachers in the country. The law is available in the Congress' website here.

requirement to *all* teacher colleges in the country.[18]

The data suggest that the NLCD screening policy was successful at reducing the fraction of low-scoring students enrolled in teacher colleges. As illustrated in Figure 9a the threshold crossing effect at 500 points jumps markedly in 2017 when the NLCD takes place, even though the BVP policy had already been in place for six years.

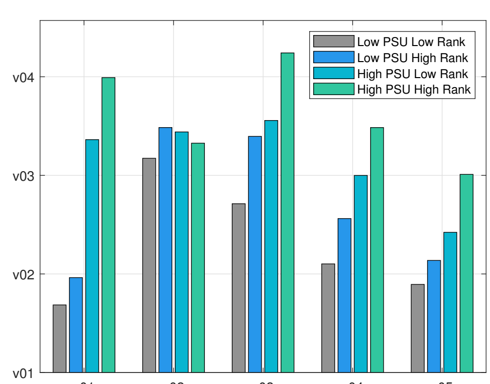
We argue that the difference between the coefficients in 2016 and 2017 gives us the NLCD effect on top of the BVP policy. Both Figure 9a and Table 5 show that the coefficient was 3.3 pp in 2016 and jumped to 6.1 pp in 2017 (and 5.3pp in 2018). These parameters suggest that the NLCD policy halved the fraction of low scorers who enrolled at teacher colleges. As a robustness (placebo) test, the results in Figures 9b 9c and their regression analogs in Table 5 show no effects at the higher scoring cutoffs of 600 and 700 points.

4.3.1 *Simulating the NLCD Screening Rule Back In Time*

We end this section simulating the policy rule of 2017 backward, spanning years 2007 through 2016, to estimate partial equilibrium effects back in time. In Figure 10, we compare labor market outcomes of prospective teachers in 2011-2016. The results show that students who would have been rejected by the 2017 screening policy performed worse in a host of labor outcomes measures.

For instance, only 10% of students who would have been rejected by the policy were likely to have a satisfactory performance in the Exit Exam, which is a 74% lower than the probability for the average accepted student. Similarly, a 29% graduated on time (within 6 years after enrollment), which is 10% lower than the average student accepted. Moreover, only 24% of rejected students working as teachers after 7 years, and only 64% of them worked in schools with high value added (as measured by average standardized exams residualized from socioeconomic status and family background variables); whereas the average accepted student had a 38% chance of teaching after 7 years, and 75% of them in high value-added schools. Finally, only 12% of rejected students were classified as good teachers by the portfolio examination, half as likely as the accepted teachers.

Figure 10: Simulation on performance outcomes



Note: Figure **??** shows the labor outcomes for each group of students enrolled in pedagogy from 2007 - 2016.

[18]The requirements for the screening policy affects admissions to all teacher colleges and are designed to be implemented gradually. During the first six years (2017-2022), the screening policy (P17) requires students to either achieve an entrance exam score above the 50th percentile of the distribution when averaging math and language or, alternatively, students can also avoid the screening rule if their high school GPA is above the 70th percentile within their high school graduating cohort.

5 Towards Data-Driven Screening Policies

We now turn to assessing whether the use of data-driven algorithms may enhance the screening procedures planned for the future by the MINEDUC policymakers.

The future screening procedures. The NLCD policy states two screening procedures for the future, one for the admissions in years 2023 through 2025 (which we call P23) and another from 2026 onward (which we label P26).

According to P23, all applicants to teaching colleges must either achieve an average entrance exam score above 525 points, or have a high school GPA above the 80th percentile of their cohort. Alternatively, applicants with less than 525 points but more than 500 points might also enroll if they also have a high school GPA above the 60th percentile.

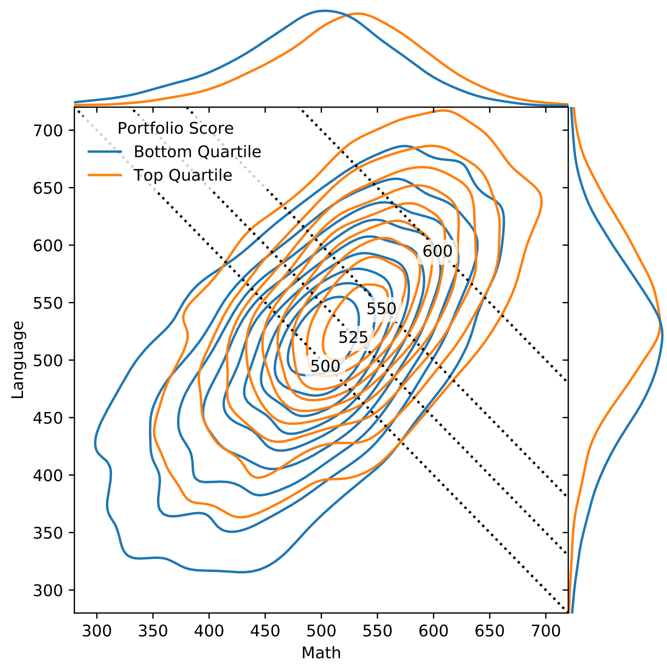
From 2026 onward (P26), the NLCD requires applicants to have either entrance exam scores above 550 points, or belong to the top 10% of their cohort high school GPA. If a student's GPA is in the top 30% *and* her average score is at least 500, then that student may also enroll at teaching colleges.[19]

Using machine learning tools. The screening rules aim to recruit promising students into the teaching profession, and therefore their effectiveness critically hinges on their predictive capacity. The availability of rich individual-level data and the low cost of prediction deems the use of machine learning (ML) algorithms as a natural way of complementing and augmenting the teacher-selection procedure, just as in other recruitment use cases (Agrawal et al., 2018). In our setting, these methods could serve as tools to augment the current cutoff-based selection rules.

Figure 11 motivates the promise of a data driven approach. The figure plots the bivariate density of math and language scores for teachers in the top (in orange) and bottom (in blue) quartiles of the portfolio evaluation, which we use here as a proxy for teacher performance. The graph shows a substantial overlap in the level curves of the score densities for both teachers categorized as top or low performers. This *prima facie* evidence suggests that there is information in the data that would be lost if classifiers rely on single-dimensional cutoff rules (Elizondo et al., 2012).

[19] 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 the admission process each year.

Figure 11: Level-curves for math and language scores, by portfolio evaluations



Note: This figure shows the level curves of the bivariate density of math and language test scores, conditional on groups defined by portfolio scores. Blue lines correspond to teachers that performed in the bottom quartile of the portfolio-score distribution, while orange lines depict those that performed in the top quartile of this distribution. Nine level curves divide each distribution in homogeneous segments that contain 10% of the data. Relevant score-cutoffs are plotted as referential dotted lines.

Procedures and Data. We show how simple classifiers can outperform MINEDUC's screening procedures. Our first exercise avoids fitting complex classifiers and is limited to families of algorithms that are close to the cutoff rules used by MINEDUC: shallow classification trees that only use math and language scores as features.

Our decision-tree classifiers use the applicants' individual math and language test scores, along with their arithmetic and geometric means, as training features.[20] Our full sample consists of about roughly 50K observations (N=49,274), corresponding to all teachers who took their college entrance exam and exit exams in 2004-2018.

We randomly split this sample into a training set (70% of sample) and testing set (the remaining 30%). Our target variable is an indicator that equals 1 when a teacher scores at least -0.5 in both standardized exit exams. We picked this threshold to create fairly well-balanced groups, with 57% of the data belonging to the positive class.[21]

[20]The geometric mean $\sqrt{Math \cdot Language}$ is a measure of the complementarity between these two scores and its relative importance in the screening process. For example, a high-math-score applicant might not be a good teacher if she performs poorly in the language test and is not able to communicate well when lecturing. This applicant could have a high arithmetic, but a low geometric mean and, in principle, would be more likely to be screened out in our procedure *vis a vis* MINEDUC's screening rules. We report analogous estimation procedures, excluding the geometric average, in the *online appendix*.

[21]We show in the *online appendix* that results are qualitatively equivalent when either changing these thresholds for the exit exams, or when repeating the procedure using other on-the-job evaluations as target variables.

Results. Figure 12 shows that our classification tree outperforms all the different government recruiting policies.

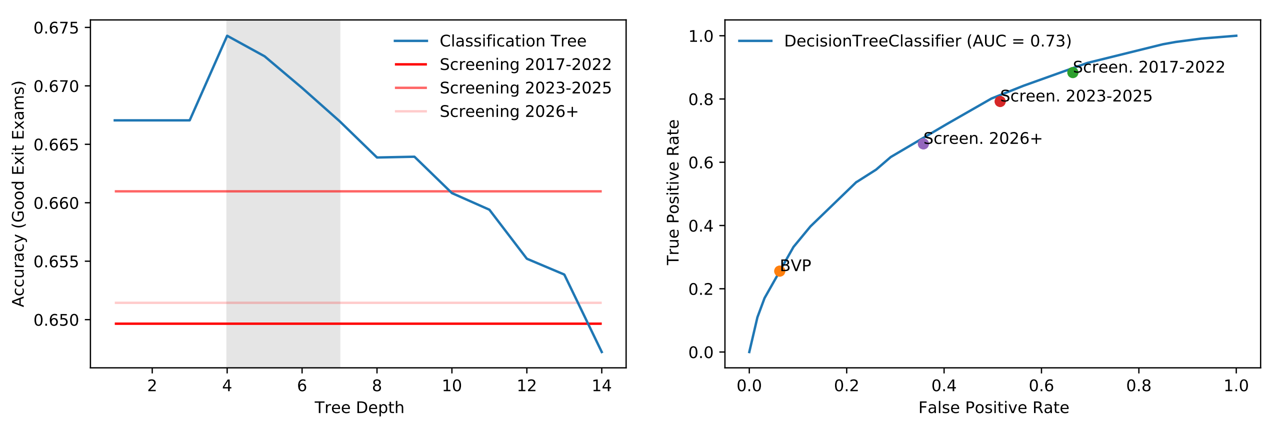
The figure in the left panel displays in blue the testing-accuracy of our tree (in the y-axis), trained with 1 to 14 depth hyperparameters (in the x-axis). The horizontal red lines correspond to the full-sample accuracy achieved by the different screening rules enacted by MINEDUC.[22] We include a shaded area that denotes the depth interval going from four (which maximizes accuracy) to seven (which achieves maximum sensitivity).

Our shallow classification tree achieves a higher accuracy compared to all NLCD screening policies by between 1.7-2.8 percentage points (a 2.5-4.2% increase). It also outperforms the BVP policy (not depicted in the left graph) by over 12 percentage points (a 22% increase).

The graph in the right panel depicts the ROC curve of our trained classification tree with four levels of depth, plotting the true positive and the false positive rates in the testing-sample. The annotated dots correspond to the location of MINEDUC's policies in this space.

The results show that our classifier achieves a higher performance than the government recruitment policies, because its ROC curve lies above all of the MINEDUC's alternatives. In addition, the area under the ROC curve (AUC) is 73% of the unit square, which is higher than the standard for predicting behavioral outcomes (Chalfin et al., 2016).

Figure 12: Classification Tree performance, compared to MINEDUC policies



Note: [Left Panel] This figure shows the testing-accuracy of our classification tree, trained with different depths. Red lines correspond to the full-sample accuracy achieved by the different screening rules enacted by MINEDUC (considering GPA rankings as well). The shaded area denotes the interval going from the depth that achieves maximum accuracy to the depth that achieves maximum sensitivity (not shown in this graph). The accuracy of the BVP policy is considerably lower (55%) and is excluded from the figure for aesthetic purposes. Another figure comparing our classifier with the BVP policy is presented in the *online appendix*. [Right Panel] This figure shows the True Positive and False Positive rates along the ROC curve of our trained classification tree, with 4 levels of depth, in the testing-sample. The annotated dots correspond to the location of MINEDUC's policies in this space. All policies are below the ROC curve, which achieves an area of 73% of the unit square.

These indicators of good performance are conservative because we kept the procedure simple. First, we are restricting our optimization space to shallow classification trees, which are low-complexity and easily interpretable algorithms.[23] Second, we do not rely on additional features, such as the high school GPA that is used in the policy screening rules. We relax these procedures below.

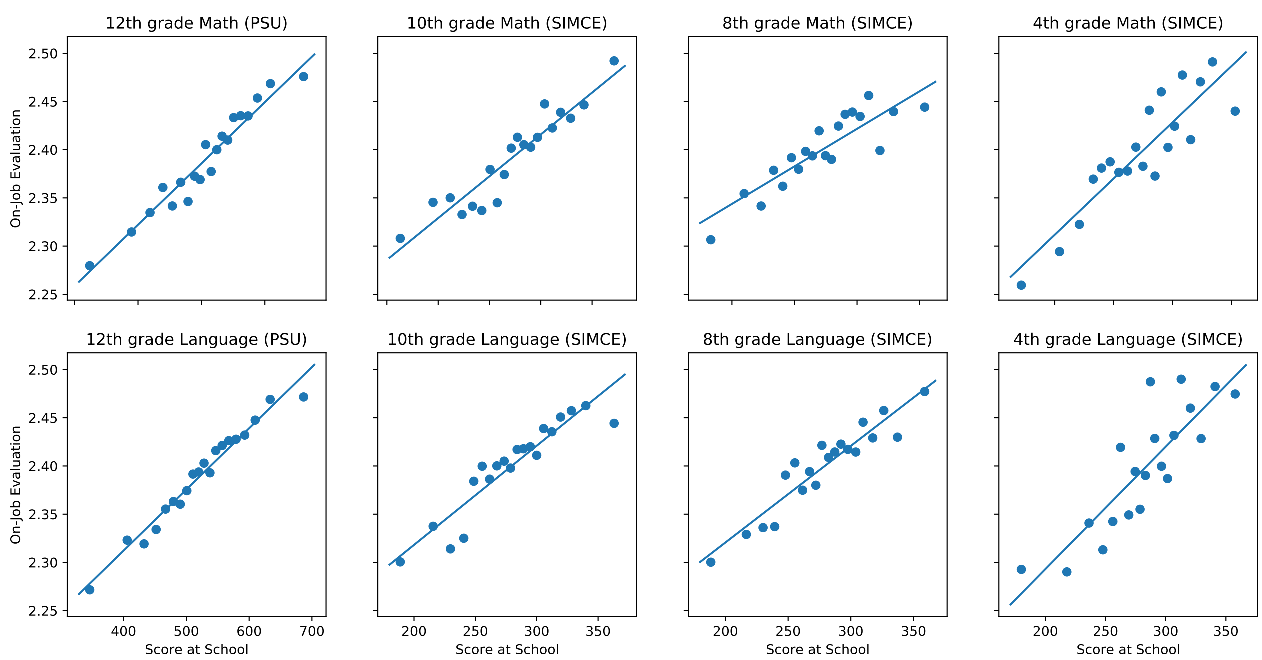
[22] The accuracy of the BVP policy is considerably lower (55%) and is excluded from the figure for aesthetic purposes.

[23] We depict the resulting optimal decision tree in the *online appendix*.

Including additional pre-college characteristics. Using the two test scores with a 4-step, if-else resulting rule[24] makes this previous classifier a useful policy tool, especially when it comes to having applicants understand the screening mechanism.

The algorithm performance might be enhanced by including additional features that also determine future performance as a teacher. We explore if other pre-college characteristics, unaffected by the screening policy, are highly correlated with on-the-job performance. We include teacher's own tests scores on standardized exams taken back in 4th, 8th, 10th and 12th grades and correlate them with their teacher evaluations as adults in Figure 13. The graphs in Figure 13 show a high correlation between teacher's own earlier scores and their performance later on, suggesting that additional features could indeed increase the predictive power of our classification algorithms.

Figure 13: School Test Scores Correlate with On-Job Performance



Note: This figure shows binned scatter plots and fitted regression lines. The dependent variable is the raw on-the-job evaluation score (portfolio). The independent variables are the their own math and language test scores on standardized exams (called SIMCE exams) taken in 4th, 8th, 10th grades, and the PSU exam taken at the end of 12th grade.

A More Complex Classifier. We now train more complex classifier using a richer feature-space to enhance the accuracy of our data-driven screening rule. Besides features, our previous algorithm might also be too constrained in terms of flexibility,[25] so we train a 16-input Multi-Layer Perceptron (MLP) with 3 hidden layers. We also use random dropout to prevent overfitting (Srivastava et al., 2014), as the number of parameters in our network (1666) is large relative to our data (19,271 observations without missing features, 70% of which are used in training). We provide further details on the features, training process, and the network architecture in the *online appendix*.

Table 6 shows that our MLP outperforms our previous simple classification tree

[24]Our best classification tree in terms of out of sample accuracy requires 4 levels of depth, which translates into 4 "if-else" branching points when classifying a data point.

[25]The family of 4-layer binary classification trees is limited in its capacity to classify datasets that are not "purely" separable in 4 steps.

and all the government recruiting policies.[26] The table displays key performance metrics (in columns) for classifiers (in rows), starting with the MLP and decision tree classifiers, and following with the MINEDUC's recruiting and screening policies: BVP, NLCD and the future NLCD policies P23 and P26.

The MLP exhibits two percentage points of additional accuracy than our previous classification tree, and beats the F1-score of the latter by two percentage points as well. Sensitivity and Precision are individually outperformed by other "extreme" rules, such as the 2017-2022 screening rule, which most applicants comply with (and therefore high sensitivity comes at the cost of a high false-positive rate), or BVP, which screens out the vast majority of applicants (and therefore high precision comes at the cost of a low sensitivity). Overall, the estimated MLP would be preferred in our setting, and it would be followed by our estimated classification tree.

Table 6: Performance comparison among different classifiers

Classifier	**Accuracy**	**Sensitivity**	**Precision**	**F1-Score**
Multi-Layer Perceptron	0.69	0.83	0.70	0.76
Classification Tree	0.67	0.80	0.68	0.74
Beca Vocacion de Profesor	0.55	0.26	0.85	0.39
Screening 2017-2022 (NLCD)	0.65	0.88	0.64	0.74
Screening 2023-2025 (P23)	0.66	0.79	0.67	0.73
Screening 2026+ (P26)	0.65	0.66	0.71	0.68

Note: This table shows various performance metrics for different classifiers. All metrics are computed in the testing data, and use as target variable an indicator that equals 1 when both standardized exit exams are at least -0.5. Accuracy: fraction of correctly classified observations. Sensitivity: true-positive rate (fraction of "good" teachers that were predicted to be "good"). Precision: positive predictive value (fraction of predicted "good" teachers that are effectively "good"). F1-Score: harmonic mean between sensitivity and precision.

Taken together, the findings of this section support the use of machine learning methods as a promising way aiding screening and recruitment policies. Importantly, we find that these methods can outperform traditional cutoff-based mechanisms without sacrificing interpretability nor relying on complicated sets of input features.

6 Conclusions

In this paper, we put together historical datasets with administrative records on the population of teachers in a middle income country to show that (i) teacher effectiveness might be predictable thanks to better data availability; (ii) teacher recruitment policies can bring better students to teacher colleges, and they will work later in schools; (iii) it remains challenging to attract very top students to the profession, (iii) combined with the development of improved algorithms, better data are lowering the cost of making accurate predictions and might improve hiring in the teacher labor markets.

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 nec-

[26] Other classifiers with richer features and more complicated architectures might outperform our MLP as well.

essary 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.

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 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 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. 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 (flexible) higher wages (Biasi, 2021). This setup should ease 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 related work should study the equilibrium effects of these policies (in the vein of Tincani, 2021), as they will likely affect the dynamic 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.

Alvarado, Macarena, Fabian Duarte, and Christopher Neilson, "Getting Better Students to Become Teachers: the Effects of Targeting Tuition Subsidies," *Working Paper, Centro de Estudios MINEDUC, Chile*, 2011.

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

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.

_ , "The Labor Market for Teachers under Different Pay Schemes," *American Economic Journal: Economic Policy*, August 2021, *13* (3), 63–102.

_ **and Heather Sarsons**, "Flexible Wages, Bargaining, and the Gender Gap*," *The Quarterly Journal of Economics*, 08 2021, *137* (1), 215–266.

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.

Cattaneo, Matias D, Michael Jansson, and Xinwei Ma, "Manipulation testing based on density discontinuity," *The Stata Journal*, 2018, *18* (1), 234–261.

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.

Elizondo, David A., Ralph Birkenhead, Matias Gamez, Noelia Garcia, and Esteban Alfaro, "Linear separability and classification complexity," *Expert Systems with Applications*, 2012, *39* (9), 7796 – 7807.

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

— **and Maria Lombardi**, "Skills and Selection into Teaching: Evidence from Latin America," Working Paper, Mimeo. 2020.

Evans, David, Feo Yuan, and Deon Filmer, "Are Teachers in Africa Poorly Paid? Evidence from 15 Countries," CGD Working Paper Series 538, Center for Global Development August 2020.

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

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

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

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

Gregorio, Soledad De and Christopher Neilson, "Teacher Preparation and Student Learning: The Ensena Chile Case," Working Paper, Mimeo 2020.

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.

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

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

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

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

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

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

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

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

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

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

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

Lombardi, Maria, "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.

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

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

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

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

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

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

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

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

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

— , *OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile*, Paris: Organisation for Economic Co-operation and Development, 2013.

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 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.

See, Beng Huat, Rebecca Morris, Stephen Gorard, Dimitra Kokotsaki, and Sophia Abdi, "Teacher Recruitment and Retention: A Critical Review of International Evidence of Most Promising Interventions," *Education Sciences*, 2020, *10* (10).

Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, 2014, *15* (56), 1929–1958.

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, 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).

Tincani, Michela M., "Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from Chile," *Quantitative Economics*, January 2021, *12* (1), 173–216.

Tincani, Michela Maria, "School Vouchers and the Joint Sorting of Students and Teachers," *Human Capital and Economic Opportunity Global Working Group (HCEO) Working Paper No. 2014-012*, 2014.

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.