Online Appendix
Screening and Recruiting Talent At Teacher Colleges*

Sebastian Gallegos † Christopher A. Neilson ‡ Franco Calle §

This version: January 10, 2020

*Usual disclaimers apply.

†Inter-American Development Bank, alvarogal@iadb.org
‡Princeton University, cneilson@princeton.edu
§Princeton University, fcalle@princeton.edu

Contents

List of Figures

List of Tables

A Teacher Exit Exam

A.1 Description and Implementation

The teacher exit exam, called 'INICIA', consists of a set of tests taken by newly graduated teachers, implemented for the first time in year 2009 and up to 2015.[1] The exam intends to measure four dimensions: (1) disciplinary knowledge (e.g. math knowledge for math teachers); (2) pedagogical knowledge (intended to measure if test takers can explain concepts in a coherent way); (3) writing skills, and (4) capacity to use ICT (information and communication technologies) for teaching purposes. In 2016 the ministry of education administered 'Diagnostica' which also evaluates disciplinary and pedagogical skills and is taken the year before graduation in different universities.

INICIA and Diagnostica's main objective is to assess the qualification of recent teacher graduates. The information produced by the exit exams is thought to be useful for the institutions training teachers, policy makers and the test-takers themselves, although there are no associated consequences to its results.[2] Results are published at the institution level, with individual-level information remaining confidential. The exam's application was gradually expanded by year and by the level at which teachers specialize (i.e. pre-school, primary and secondary), as summarized in Table 1.[3]

[1] Typically, students in their last semester of class, or just graduated students taking the exam before getting a job.

[2] This may change in the near future. During 2015, the Ministry of Education sent a bill to the Congress in order to make the INICIA test mandatory and to establish minimum performance levels to be allowed to teach at least in the public sector.

[3] In 2013 the exam was not applied.

Table 1: Teacher Exit Exam: Tests Implemented by Year and Teacher Specialization Level

		Year						
Level	Test	2009	2010	2011	2012	2014	2015	2016
Pre-school	Disciplinary	✓	✓	✓	✓	✓	✓	✓
	Pedagogical		✓	✓	✓	✓	✓	✓
	Writing	✓	✓	✓	✓			
	ICT							
Primary	Disciplinary	✓	✓	✓	✓	✓	✓	✓
	Pedagogical	✓	✓	✓	✓	✓	✓	✓
	Writing	✓	✓	✓	✓			
	ICT	✓	✓	✓				
Secondary	Disciplinary				✓	✓	✓	✓
	Pedagogical				✓	✓	✓	✓
	Writing				✓			
	ICT							

Notes: 'Disciplinary' stands for the test measuring disciplinary knowledge; 'Writing' stands for the writing skills test; 'Pedagogical' stands for the Pedagogical knowledge test; 'ICT' stands for the test measuring the information and communications technology skills. Source: MINEDUC (2012).

The Inicia exam is voluntary. Formally, the Ministry of Education invites graduate institutions that train teachers (i.e. institutions offering teacher or education degrees) to participate in the INICIA exam every year. In the case of Diagnostica, the exam is mandatory and is administered to all students of pedagogy in certified institutions of education. Table 2 summarizes the number of institutions invited, and those that participated. From years 2009 to 2012, around 80% of the invited institutions participated, which means that at least some of its graduates took the test. Institutions can encourage their graduates to participate, but can not force them to do so.

Table 3 summarizes the number of potential test-takers, the ones that sign-up and those that take at least one test, by year. Every INICIA test before 2012 was held in December of

each year, which coincides with the end of the academic year in Chile. Due to administrative issues, the 2012 INICIA test was held in April of 2013. This delay seems to be the reason behind the low take-up of that year's test (see Table 3). By that time of the year, most new teachers would be working, because the academic year starts in March. Also, it is likely that graduates lose the connection with their universities after a while. After this episode, the Ministry of Education decided to postpone the application of the 2013 INICIA, supposed to be held in December 2013, to December of 2014, combining evaluations 2013 and 2014 into a single sitting.

Table 2: Teacher Exit Exam: Invited and Participating Institutions by Year

	Application	Number of Institutions		Participation
Year	Date	Invited	Participating	Percentage
2009	Dec. 2009	54	43	80%
2010	Dec. 2010	56	43	77%
2011	Dec. 2011	59	49	83%
2012	Apr. 2013	58	50	86%
2014	Apr. 2014		50	-
2015	Dec. 2015		50	-
2016*	Apr. 2016		50	-

Notes: Invited institutions correspond to those that train primary school teachers (every year), pre-school teachers (years 2009-2012 and 2016) and secondary school teachers (year 2012 and 2016). Participating institutions are the ones for which at least one of their graduates takes one or more of the tests described in Table 1. Participation percentage displays the number of participating institutions as a percentage of the number of invited institutions. * In 2016 the corresponding exit exam was Diagnostica and was mandatory. Source: MINEDUC (2012).

A.2 Teacher Exit Exam Results and PSU Scores

Institutional Reports. The Ministry of Education publishes each year a presentation with the INICIA exam results.[4] According to these institutional reports, the results achieved by the education graduates are below what is needed to perform adequately as a teacher.

[4]For years 2008 to 2010, results were mainly published as the percentage of correct answers achieved in each test, without a statement on what was considered a good outcome. For the 2011 and later exams, the MINEDUC implemented three categories to classify test-takers according to their performance, based on the knowledge and skills necessary to begin their career as a classroom teachers: Outstanding, Acceptable and Unsatisfactory.

Table 3: Teacher Exit Exam: Test-Takers by Year

	Number of Test-Takers			
Year	Potential	Signed-up	Participated	Take-up
2008	5,250	3,006	1,994	38%
2009	7,979	4,527	3,223	40%
2010	8,594	4,681	3,616	42%
2011	8,069	4,874	3,271	41%
2012	10,351	2,443	1,443	14%
2014	15,013	714	682	4%
2015	14,472	1,993	1,916	13%
2016*	20,215	17,971	12,741	63%

Notes: Potential Test-Takers correspond to the number of graduates from previous year. Those that sign-up to take the test are displayed in column 2. The number individuals that took at least one test described in Table 1 is shown in column 3. Column 4 presents the number of actual test-takers as a percentage of the potential test-takers. * In 2016 the corresponding exit exam was Diagnostica and was mandatory. Source: MINEDUC (2012).

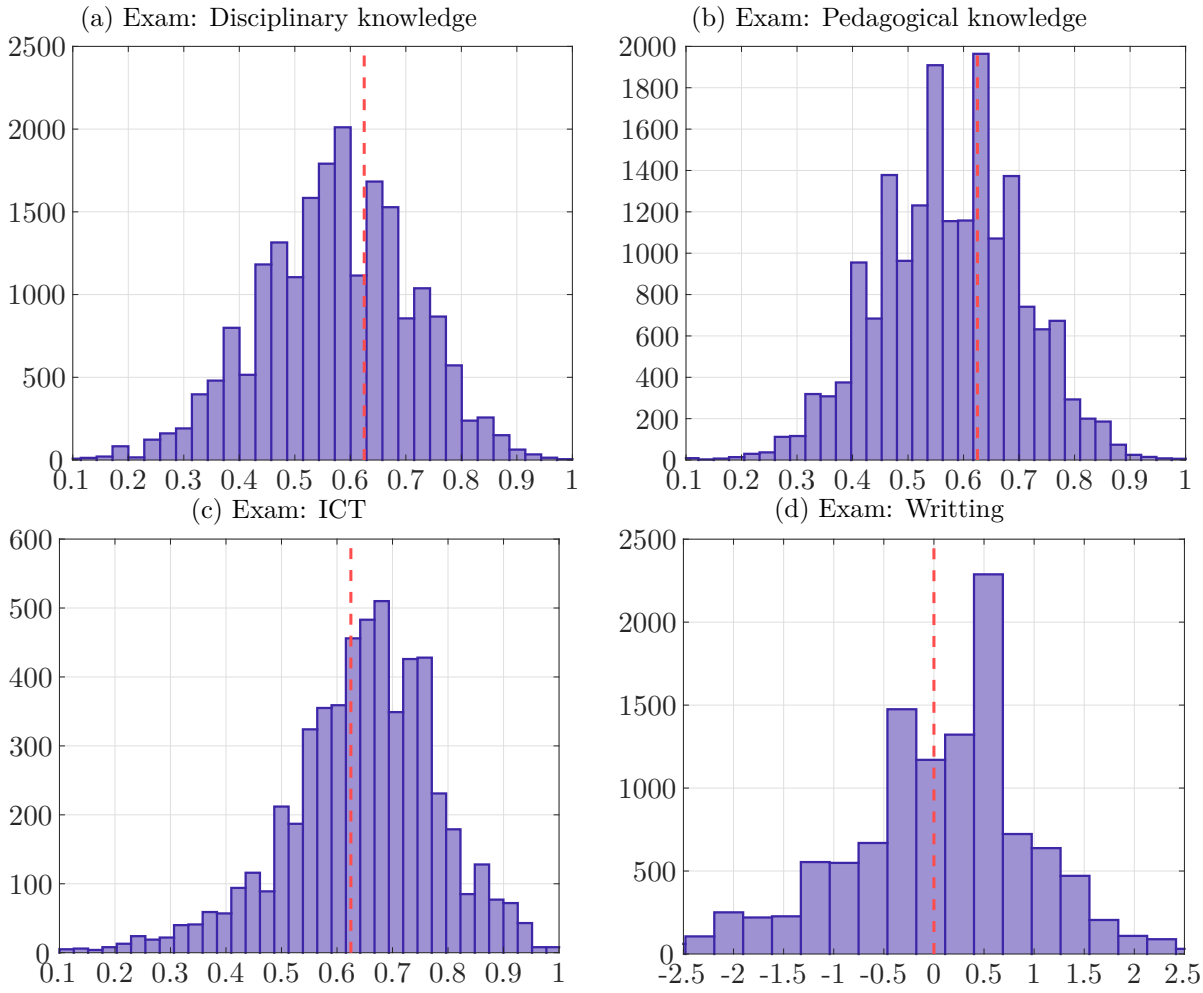
More than 60% of the test-takers that graduated as primary teachers fall in the 'unsatisfactory' category for the disciplinary tests in 2011 and 2012. The percentage is approximately 40% for the pedagogical test. For secondary teachers, the disciplinary tests by subject show the worse results in Mathematics, Biology, Physics and Chemistry, where about 70% of the test-takers fall in the 'unsatisfactory' category.

Microdata. The Ministry of Education provided us information from 2009 to 2015 on the INICIA exam, at the individual level and 2016 data on Diagnostica exam. We have microdata for more than 16K teachers with INICIA scores in at least one test and arround 13k teachers evaluated in Diagnostica. Table 4 provides summary statistics for the four available tests. The first three rows report the percent of correct answers for the disciplinary, pedagogical and the ICT tests.[5] The last row shows the scores in the standardized writing test.

Figure 1 shows histograms for the four tests, where the vertical dashed line indicates the

[5]The difference in the samples is explained by the fact that the Pedagogical test was not held in 2009, and the ICT test was not applied in 2012.

Figure 1: Exit Exams Histograms



Note: All four figures use information for teachers that took the respective tests from the INICIA exit exam between years 2009 and 2014. The dashed red line indicates the cutoff above which the performance in each test is considered 'acceptable'. These cutoffs are 0.61, 0.61, 0.65 and -0.09 for the Disciplinary, Pedagogical, ICT and Writing tests, respectively. Cutoff values vary slightly over the years, so they should be interpreted as proxies. The figures consider all test-takers with valid scores in the Disciplinary test (Figure 1(a), $N = 12,477$), the Pedagogical test (Figure 1(b), $N = 8,943$), the ICT test (Figure 1(c), $N = 6,249$) and the Writing test (Figure 1(d), $N = 10,665$).

Table 4: Exit Exam Summary statistics

Variable	Mean	Std. Dev.	Min	Max	N	Corr(PSU)
% of Correct Answers in:						
Disciplinary Test	0.57	0.14	0	1	20224	0.53
Pedagogical Test	0.58	0.13	0.05	1	18025	0.51
ICT Test	0.65	0.14	0.1	1	5517	0.51
Writing Test*	0.02	0.99	-6.65	3.26	11300	0.28

Notes: the last column displays the displays Spearman's rank correlations for each variable and PSU scores. The percentage of correct answers for the Disciplinary, Pedagogical and ICT tests has an associated a cutoff above which the performance is considered 'acceptable'. These cutoffs are .61, .61 and .65 for the Disciplinary, Pedagogical and ICT tests, respectively. These thresholds vary slightly over years, so they should be interpreted as proxies. For the writing test score, the cutoff is about -.09 SD from the mean. * Writing test is a standardized variable of the scores achieved by students by years.

cutoff above which the performance is considered acceptable.[6] Test-takers perform poorly: in the Disciplinary test, 62% of the test takers are *below* the threshold. For the Pedagogical test, the percentage is 58%. For the ICT and Writing tests, 39% and 42% of the test-takers have scores below acceptable.

Exit exam test results are strongly correlated with PSU scores, as suggested by the Spearman's rank correlations in the last column in Table 4 and nonparametric plots of the bivariate relation in Figure 2. The positive correlation ocurs in the whole spectrum of the PSU score according to the figures.

Bear in mind that the PSU test is administered right before beginning higher education studies, and the exit exams are administered right after completing their degree.

B Teacher Evaluation in the Public Sector

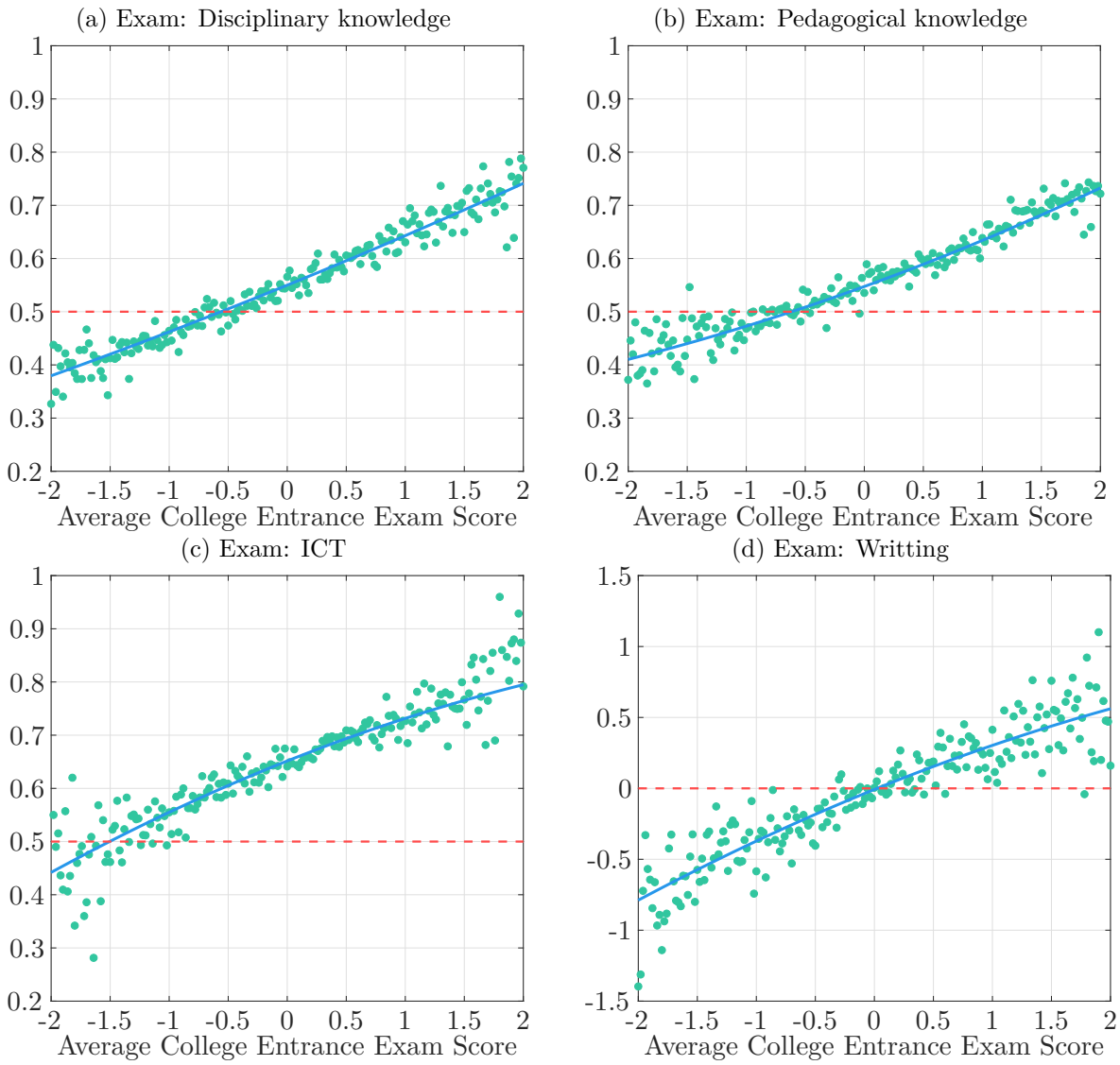
B.1 Teacher Evaluation Description and Implementation

The Teacher's Public Evaluation System[7] (*Evaluación Docente* in Spanish, or ED onwards) is a mandatory assessment for all classroom teachers working in the public sector in Chile.

[6] This cutoffs vary slightly by year, so they should be considered general guidelines and not absolute thresholds.

[7] For more details, see the institutional website www.docentemas.cl.

Figure 2: Exit Exams vs PSU scores



Note: All four figures use information of teachers that took the respective Inicia exit exam test between years 2009 and 2014. PSU scores have a mean of 500 points and a standard deviation (SD) of 100 points, so each plot's x-axis shows ± 2 SD. Solid blue lines correspond to fitted values from local linear regressions using a rectangular kernel with a bandwidth of 10 PSU points, with 95% confidence intervals plotted in gray. The open circles plot the average values of each variable within five points of the PSU score. The dashed red line indicates the cutoff above which the performance in each test is considered acceptable; these cutoffs are .61, .61 and .65 for the Disciplinary, Pedagogical and ICT tests, respectively and -.09 for the Writing test. The cutoffs vary +-slightly over years, so they should be interpreted as proxies. The Figures consider all test-takers with valid scores in the PSU and the Disciplinary test (Figure 2(a), $N = 11,060$), the Pedagogical test (Figure 2(b), $N = 7,447$), the ICT test (Figure 2(c), $N = 5,795$) and the Writing test (Figure 2(d), $N = 9,908$).

The ED declared objective is 'to strengthen the teaching profession and the quality of education'. The assessment is composed by four components, with different weights: (i) a self-evaluation questionnaire (10%); (ii) a third-party reference report, filled by the school principal or supervisor (10%); (iii) one peer review (20%), and a *teacher performance portfolio* (60%). The portfolio component aims to collect direct evidence on teaching skills, pedagogical decisions and classroom practice. It includes two modules. In the first module, teachers plan a class defining its contents and related assessments. They are also asked questions about teaching practices. The second module consists in a videotaped class followed by a questionnaire on the students behavior and understanding, and the teacher's own performance.

The ED assigns a weighted score for each teacher using the components (i) to (iv) above. Then, the score is used to classify each teacher performance in one of four categories: unsatisfactory, basic, competent or outstanding. As opposed to the INICIA exit exam, the ED has consequences associated to performance. Teachers classified in the 'competent' or 'outstanding' categories can opt to receive a monetary bonus. Teachers classified in the unsatisfactory level need to retake the ED. If they remain in the unsatisfactory category after three times, they must leave their schools and can not teach again.

The ED has been implemented gradually since 2004 according to the level at which teachers specialize (pre-school, primary, secondary).[8] Table 5 shows its year-level coverage for ten years 2004 to 2016.

[8]There are also other levels that have been incorporated to the teacher evaluation, like special education and education for adults, but we focus on preschool, primary and secondary levels in our analysis.

Table 5: Teacher Evaluation Implementation by Year and Level Taught

	Level		
Year	Preschool	Primary	Secondary
2004		✓	
2005		✓	✓
2006		✓	✓
2007		✓	✓
2008	✓	✓	✓
2009	✓	✓	✓
2010	✓	✓	✓
2011	✓	✓	✓
2012	✓	✓	✓
2013	✓	✓	✓
2014	✓	✓	✓
2015	✓	✓	✓
2016	✓	✓	✓
2017	✓	✓	✓

Notes: There are also other levels that have been incorporated to the teacher evaluation, like special education and education for adults, but we focus on primary and secondary levels in our analysis.

The ED has carried out more than 174207 assessments for preschool, primary and secondary teachers from 2004 to 2013. Table 6 exhibits the number of evaluations per teacher by year. The system has evaluated 101423 teachers at least once.[9] Approximately half of those teachers have been evaluated twice[10] (∼51K), and a about 35K have been evaluated more than three times.

For purposes of the analysis we will restrict the sample to teachers of primary or secondary education that were evaluated. This sample consist on 78513 teachers from the total of 101K evaluated (%77 of the total sample). Table 7 reports the first ED results per

[9]From the 101K evaluated teachers a fraction has already retired from teaching. To get a sense of the coverage regarding those working currently in the public sector, consider that in year 2016 130K classroom teachers were working in the public sector (in either the preschool, primary or secondary level) and about 101K of them (∼78%) had been evaluated at least once.

[10]All teachers are supposed to be re-evaluated every four years, which the data does not fully support; teachers first classified in the unsatisfactory or basic category should be re-evaluated the year after or two years after the first evaluation respectively.

Table 6: Number of times teachers were evaluated from 2004-2013

Year	N:1	N:2	N:3	N:4	N:5	N:6
2004	1719	0	0	0	0	0
2005	10631	34	0	0	0	0
2006	13931	255	4	0	0	0
2007	10178	208	27	0	0	0
2008	14890	1104	21	0	0	0
2009	8567	5524	25	0	0	0
2010	3873	6422	121	3	0	0
2011	3498	7274	158	9	0	0
2012	3875	10496	693	17	0	0
2013	4343	6447	3818	57	4	0
2014	4993	3536	5118	167	7	0
2015	4620	2828	3889	339	16	0
2016	5707	3229	6118	899	35	1
2017	6667	3657	4757	2080	47	2
All	101423	50744	20456	1518	65	1

Notes: The table above represent the number of tests administered each year by the number of times a teacher was evaluated until each year.

category for all the 78.5K teachers in its first column. Only a 2 percent of the teachers resulted in an 'unsatisfactory' performance; 28% were classified as 'basic', 61% as 'competent' and 9% as 'outstanding'. It also shows the maximum scored achieved by category for some years. The thresholds to be in each category vary by year.

Table 7: Teacher Evaluation Results 2004-2016

Classification	N obs	%	Max: 2004	Max: 2008	Max: 2012	Max: 2016
Outstanding	6875	8.8	3.63	3.59	3.21	3.37
Competent	48130	61.3	3.11	3.25	3	3.15
Basic	22091	28.1	2.64	2.67	2.79	2.9
Usatisfactory	1417	1.8	2	2.26	1.95	2.1
Total:	78513	100	2.84	2.94	2.74	2.88

B.2 Teacher Evaluation Results and PSU Scores

From the sample of first test taken by primary and secondary teachers we examine the correlation between ED and PSU scores. From the 78,513 teachers of primary and secondary education with ED scores about 63K (or 81%) have an available PSU score, while 14974 (or 19%) have not. As we explained in detail in the PSU Section, we collected data on the national college exam (PSU) that teachers took up to 35 years ago (from 1980 onwards). Therefore, we do not have information for the older teachers, many of whom have retired from teaching anyway. On average, the teachers with ED scores but no PSU scores were 61 years old in 2016, and a 44% of them was not teaching during year.

Table 8 shows the teacher evaluation results by availability of PSU scores. Panel A shows the results by the four categories of performance. We also have information on the overall ED score and also the portfolio component score, whose results we present in Panel B of Table 8.

Teachers with PSU scores tend to perform better in the ED. Panel A shows that they fall more in the upper two categories (competent and outstanding) and less in the lower (basic and unsatisfactory). Consistently, teachers with PSU scores also achieve higher ED scores, both overall and in the portfolio component as shown in Panel B. Differences in both Panels are significant at the 1% level.

Figure 5 shows the distributions of the overall ED scores and portfolio scores (Figure 5(a) and Figure 5(b), respectively) for teachers with and without PSU scores. The vertical dashed lines indicate the scores that separate the four categories.[11] Even though differences do not appear distinguishable to the eye, a two-sample Kolmogorov-Smirnov test rejects equality of distribution functions for each score. In any case, given the positive relationship between the ED and PSU scores that we document next, we expect teachers without PSU information to have lower PSU scores.

[11]The cutoffs vary slightly over years, so they should be interpreted as proxies.

Table 8: Teacher Evaluation Results 2004-2013, by PSU Score availability

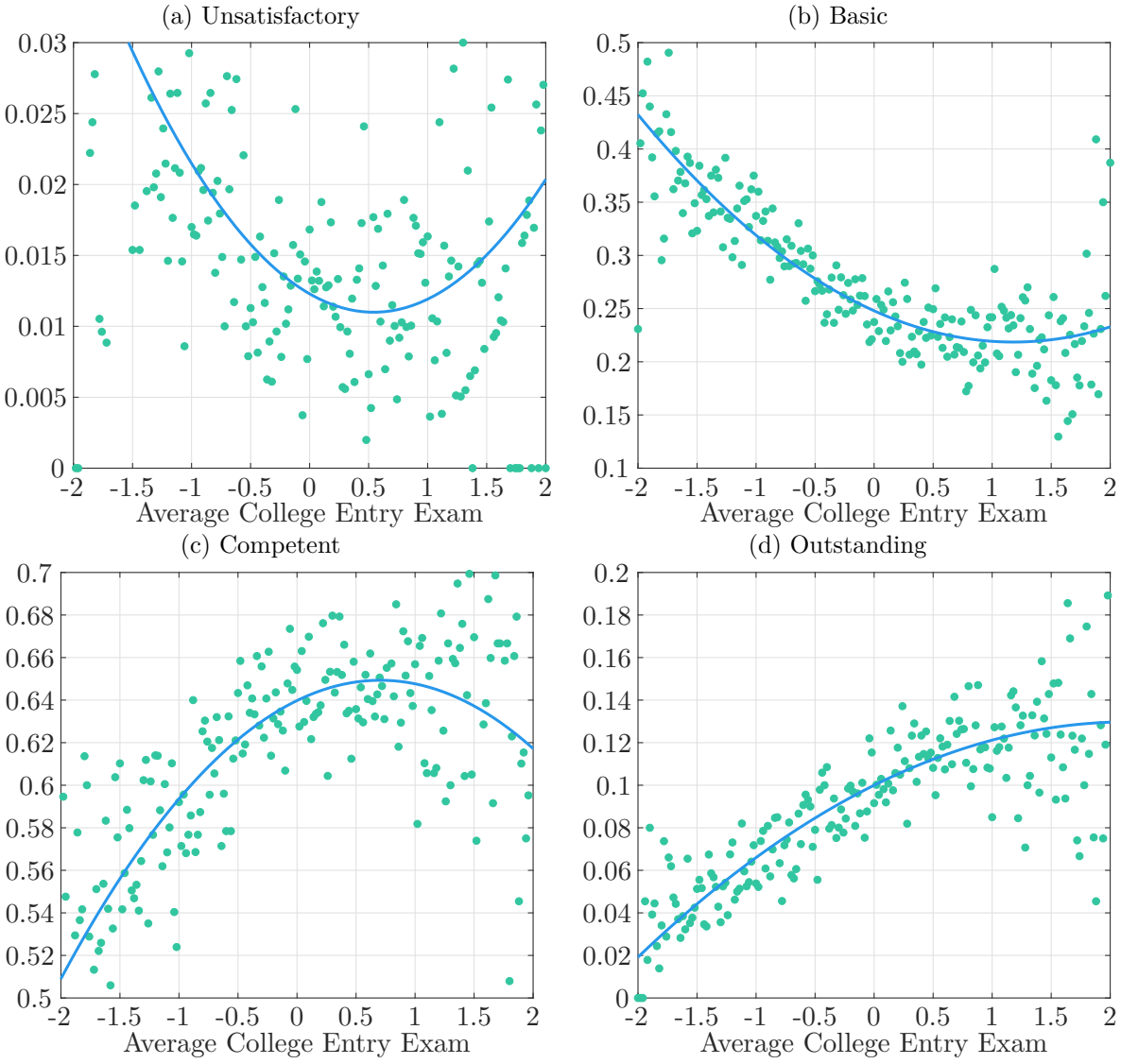
	With PSU Scores		Without PSU Scores		T-Test
Classification	N obs	%	N obs	%	Difference
Outstanding	5931	9.3	944	6.3	3 ***
Competent	39605	62.3	8525	56.9	5.4 ***
Basic	16987	26.7	5104	34.1	-7.4 ***
Unsatisfactory	1016	1.6	401	2.7	-1.1 ***
Total	63539	100	14974	100	
Score	N obs	%	N obs	%	Difference
Overall	2.63	0.28	2.57	0.29	0.06 ***
Portfolio	2.29	0.32	2.2	0.32	0.09 ***

As for the exit exams, our large set of observations allow us to accurately graph the bivariate relation between the respective ED outcomes and the PSU Scores. Figure 3 shows that there is negative relation between being evaluated as unsatisfactory (Figure 3(a)) or basic (Figure 3(b)) and PSU Scores, while there is a positive relation with being classified as competent or outstanding (Figure 3(c) and Figure 3(d), respectively). The positive relation is also clear when examining ED scores both for the overall and portfolio case as shown in Figure 4(a) and Figure 4(b).

In Table 9 we test whether this correlation persists once we control for year fixed effects, and the specialization level of teachers. The regression table reports the coefficients of separate regressions of each teacher evaluation outcome on the PSU score, expressed in terms of standard deviations. The columns add year fixed effects and teacher specialization level fixed effects. Each coefficient should be read as the change in the dependent variable given one standard deviation (SD) of increase in the PSU scores.

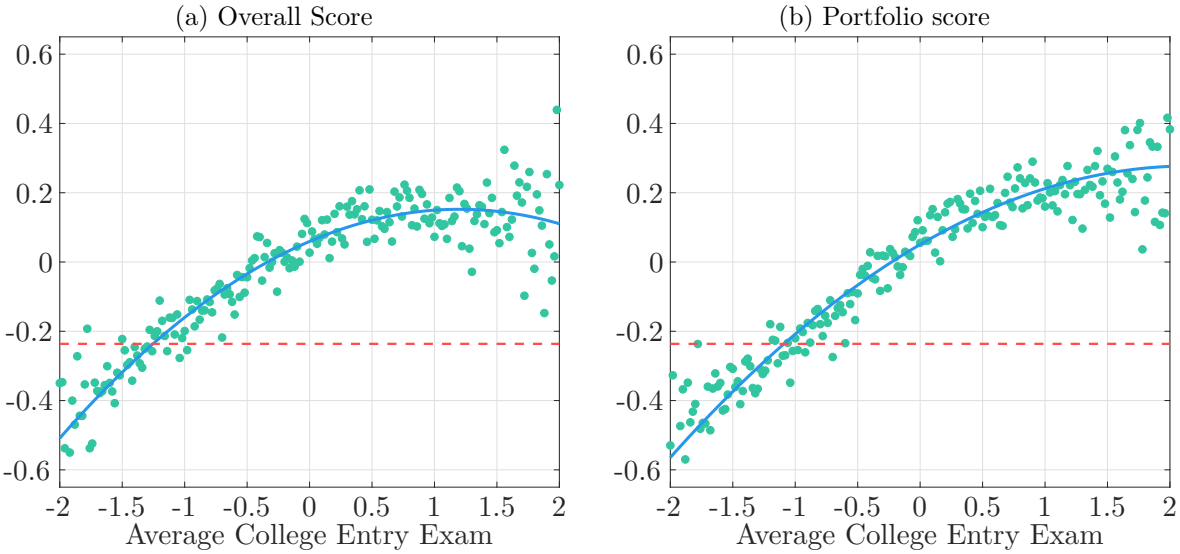
The results show that the coefficients are all significant (at 1% level) and stable across specification for the same outcome (i.e., independent of the controls added in each column). One SD increase in PSU scores is associated to an increase of approximately .7 SD standard deviations in the overall score. For the portfolio score, one SD of increase in the PSU is associated to a .6 standard deviation increase.

Figure 3: Teacher Evaluation Categories vs. PSU Scores



Note: The figures Figure 3(a), Figure 3(b), Figure 3(c) and Figure 3(d) plot the probability of being classified by the Government as unsatisfactory, basic, competent and outstanding respectively. The plots are built with 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in students enrolled in years 2004 to 2009 who graduated between 2009 and 2017.

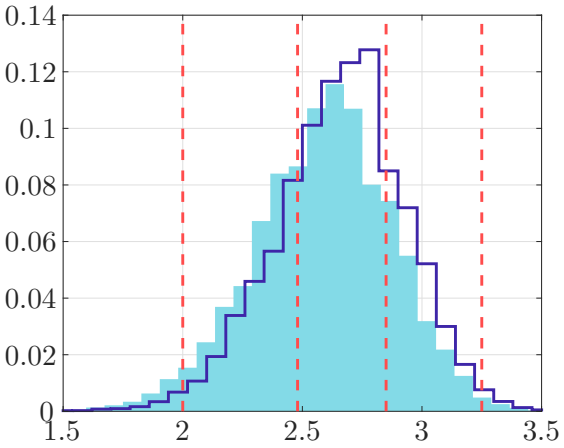
Figure 4: Teacher Evaluation Scores vs. PSU Scores



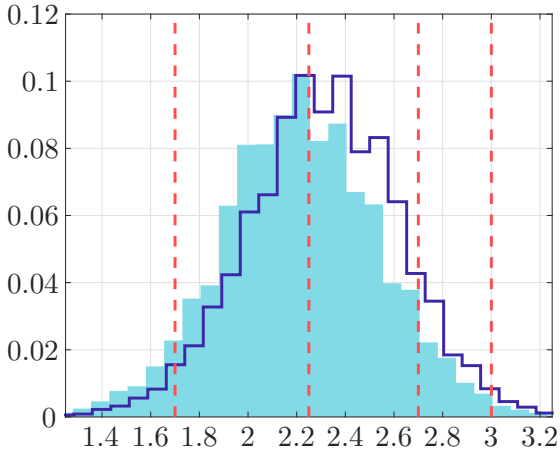
Note: Both figures use information for teachers working in the public sector that were evaluated at least once between years 2004 and 2016. The dots plot the average values of each variable within five points of the PSU score. The PSU score has a mean of 500 points and a standard deviation (SD) of 100 points, so each Figure plots data up to two SD to the left, and two SD to the right. The vertical axis the overall and portfolio scores of the teacher evaluation as a function of their PSU scores.

Figure 5: Teacher Evaluation Histograms

(a) Overall ED Score



(b) Portfolio Score



Note: Both figures use information for teachers working in the public sector that were evaluated at least once between years 2004 and 2016. Figure 5(a) and Figure 5(b) plot the histograms for the overall evaluation score and portfolio evaluation score achieved by teachers respectively. The unshaded histogram with color shows the distribution for teachers without PSU scores, meanwhile the shaded histogram with color plots the distribution for teachers without PSU scores. The vertical dashed lines indicate the scores that separate teachers into four categories of performance (unsatisfactory, basic, competent and outstanding) as explained in the text. The cutoffs vary slightly over years, so they should be interpreted as proxies.

C Teacher Wages

C.1 Teacher Wages Description

The Ministry of Education collects administrative data on all teachers working in Chile each year. Each school of the country reports their number of working teachers disaggregated at the individual level with a set of characteristics of their job, like hours of contract, level taught and the subject they teach. In 2011 the Ministry of Education also asked schools to provide the wages payed to their teachers.

About 111K (88%) have a valid wage.[12]

[12]From the 125K teachers in the private sector, 113K were reported to have positive wages. We trimmed wages below the percentile 1 and above percentile 99, which left us with 111K teachers with valid wages.

Table 9: Regressions of Teacher Evaluation Performance on PSU Scores

	(1)	(2)	(3)
Overall Score, Mean: 0, S.D: 1			
PSU Score	0.801 ***	0.725 ***	0.615 ***
S.E.	(0.042)	(0.041)	(0.041)
PSU Score2	-0.065 ***	-0.059 ***	-0.048 ***
S.E.	(0.004)	(0.104)	(0.001)
N. Obs	[63539]	[63539]	[63539]
Portfolio Score, Mean: 0, S.D: 1			
PSU Score	0.697 ***	0.589 ***	0.477 ***
S.E.	(0.041)	(0.04)	(0.04)
PSU Score2	-0.049 ***	-0.04 ***	-0.031 ***
S.E.	(0.004)	(0.101)	(0.001)
N. Obs	[63539]	[63539]	[63539]
Writing Skills, Mean: 0, S.D: 1			
PSU Score	0.194 ***	0.255 ***	0.536 ***
S.E.	(0.057)	(0.055)	(0.046)
PSU Score2	-0.015 ***	-0.022 ***	-0.049 ***
S.E.	(0.006)	(0.134)	(0.002)
N. Obs	[36771]	[36771]	[36771]
ICT Skills, Mean: 0, S.D: 1			
PSU Score	0.458 ***	0.448 ***	0.628 ***
S.E.	(0.045)	(0.044)	(0.043)
PSU Score2	-0.037 ***	-0.036 ***	-0.055 ***
S.E.	(0.004)	(0.114)	(0.002)
N. Obs	[58523]	[58523]	[58523]
Year F.E.		X	X
Controls			X

Note: Robust standard errors in parentheses. Significance levels: * p0.10, ** p0.05, *** p0.01. The regression table reports the coefficients of 12 separate regressions for different measures of productivity over PSU score, both dependent and independent variables expressed in terms of standard deviations. The columns add year fixed effects, teacher specialization level (primary and secondary levels) fixed effects and a polinumium of order two of experience. Each coefficient should be read as the change in the outcome given one standard deviation (SD) of increase in the PSU scores.

C.2 Teacher Wages and PSU Scores

About 90% from the 111K teachers with valid wages have an available PSU score. The 10% of teachers with no PSU score are older teachers, aged 44 on average, 7 years more than educators with psu available.[13] Table 10 shows that teachers with no PSU score earn higher wages, probably due to seniority. Figure 6 plots the hourly wage histogram by PSU Score availability, which confirms that for most teachers in the upper part of the distribution we do not have PSU scores.

Table 10: Hourly Wage, by PSU Score Availability

With PSU Scores	Mean	Std. Dev.	N	%	p10	p50	p90
No	19.71	6.92	10954.00	10.00	12.75	17.76	29.24
Yes	19.06	6.24	100653.00	90.00	12.97	17.64	27.29
All	19.12	6.31	111607.00	100.00	12.94	17.65	27.54

As opposed to teachers working in the public sector (39%), teachers from the voucher sector (61%) are not subject to any mandatory evaluation. However, while teachers working in public schools benefit from a special labor code, voucher teachers work under the regular, more flexible, labor code. Therefore, their wages should be associated to productivity, reflecting how the labor market values their performance. We use the market wages as a proxy for the voucher teachers quality.

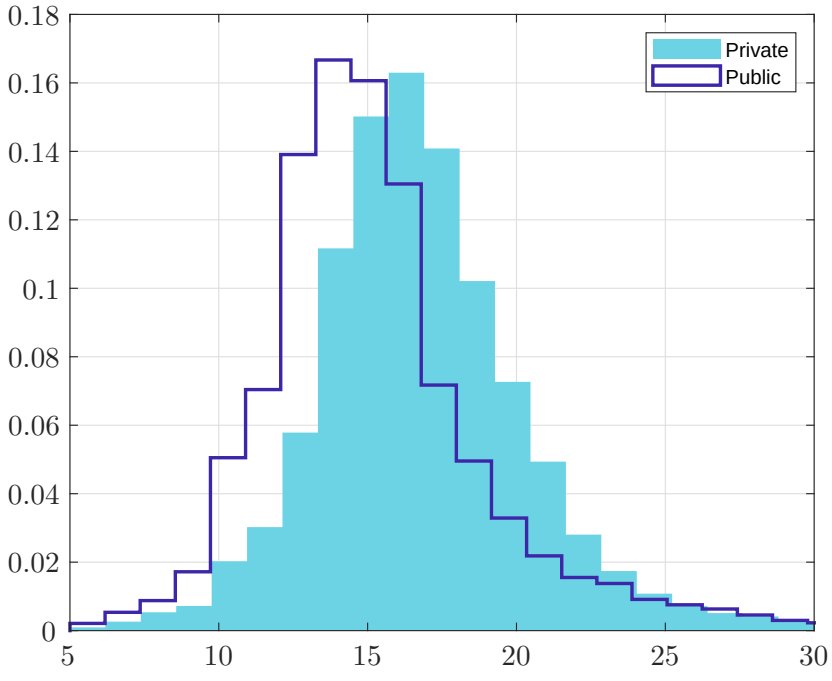
In Table 11 we report the decomposition of teachers in both public and private sector by working hours and, as can be seen, the most productive teachers working in the private sector are the ones that work full time. In contrast, in the public sector the most productive teachers are the ones that work less hours.

[13] As we explained in detail in the PSU Section, we collected data on the national college exam (PSU) that teachers took up to 35 years ago (from 1980 onwards). Therefore, we do not have information for the older teachers.

Table 11: Hourly Wage Summary Statistics

Working Hours	Mean	Std. Dev.	N	%	p10	p50	p90	Corr(PSU)
Private Sector								
$[0-30)$	18.15	6.11	19752.00	33.00	12.85	16.93	24.48	0.04
$[30-35)$	17.92	4.36	10957.00	19.00	13.33	17.25	23.60	0.07
$[35-40)$	18.07	4.17	11733.00	20.00	13.83	17.38	23.35	0.04
$[40-45]$	18.84	4.69	16711.00	28.00	14.05	18.01	24.99	0.08
All	18.31	5.13	59220.00	100.00	13.47	17.39	24.29	0.06
Public Sector								
$[0-30)$	19.56	9.00	7202.00	18.00	11.56	16.75	32.12	0.02
$[30-35)$	20.10	6.51	10231.00	26.00	12.61	18.97	29.35	-0.00
$[35-40)$	20.00	6.53	9846.00	25.00	12.61	18.76	29.01	-0.04
$[40-45]$	19.92	6.54	11265.00	29.00	12.55	18.57	28.81	0.03
All	20.19	7.40	39068.00	100.00	12.44	18.42	29.76	0.00

Figure 6: Wage distribution by sector



Note: The unshaded histogram with color shows the distribution of wages for teachers in the public sector, meanwhile the shaded histogram with color plots the distribution for wages of teachers in the private sector.

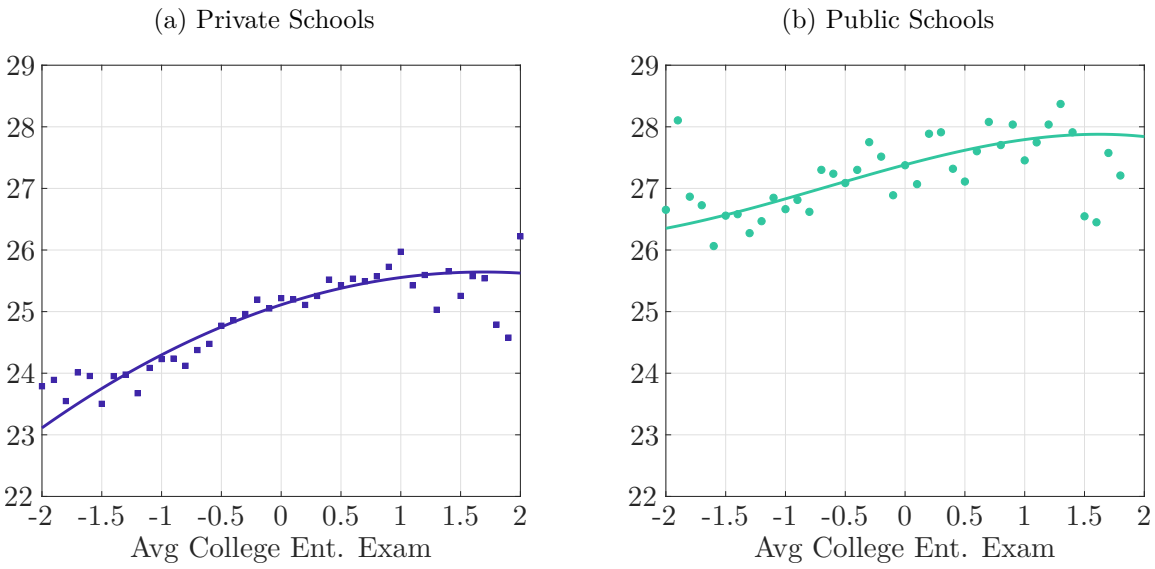
Moreover, as for previous measures of teacher quality, higher wages are associated to higher PSU scores. For our sample of 59K teachers in private schools with valid wages and PSU Scores, the Spearman correlation for those variables is positive, .06. as shown in Table 11. The correlation is higher for those teachers that work full time at the schools (.08 for the last row). However, this does not occur in the public sector, where the correlation is not clear.

Figure 11 graphs nonparametrically the bivariate relation between wages and the PSU Scores for teachers in the private and public sector. Figure 7(a) shows a concave, non monotonic relation between psu and wages in the private sector, similar to the trend with the evaluacion docente scores. And, Figure 7(b) depict the same relation but for public sector professors. The positive relation exist, however is not as strong as the one reported for the teachers in the private sector.

In Table 9 we test whether this correlation persists once we control for the teachers' age (a quadratic function of age), the specialization level of teachers and the subjects they teach. The regression table reports the coefficients of separate regressions of wages for private and public sector teachers on the PSU score, expressed in terms of standard deviations. The columns add the controls for age, teacher specialization and subject taught. Each coefficient should be read as the change in the dependent variable given one standard deviation (SD) of increase in the PSU scores. We normalized the dependent variables in this case so we can refer to the coeficient as the percentage increase of a standard deviation in the dependent variable.

The results show that the coefficients are all significant (at 1% level). The magnitude decreases as we include controls, nevertheless remains sizable. One SD increase in PSU scores is associated to a 0.4 SD in wages in the private sector and the half the increase (0.24 SD) in the public sector.

Figure 7: Teacher Wages vs. PSU Scores



Note: The two figures use information for teachers working in the voucher and public sector since 2011 until 2017, with valid wages and PSU scores. The graphs plot the wages for private sector teachers (Figure 7(a)), and public sector teachers (Figure 7(b)) as a function of their PSU Scores. The open circles the average values of each variable within five points of the PSU score. The PSU score has a mean of 500 points and a standard deviation (SD) of 100 points, so each Figure plots data up to two SD to the left, and two SD to the right.

D Employment: Working in Schools as a function of Exam Scores

D.1 Graduating Cohorts from Education Majors 1995-2013

We gathered records from students who graduated from all education majors in Chile for nine specific years between 1995 and 2013. We then combined them with administrative records on teachers who were working in schools in 2011.

Table 12 shows the number of graduates by year, with information on the fraction working in schools in 2015 and their PAA Scores. On average a 46% of 127K graduates from 1995-2013 were working in schools in 2011, and they have on average a score of 508 and the likelihood of working as teacher years after vary with the graduates' PAA scores, as Figure 8 shows.

Figure 8 shows the fraction of graduates of 2007 working in schools 2, 5, 10 years after

as a function their PAA Scores. The inverted 'U' shape of the solid line in suggests that both low and high scored graduates have a lower likelihood of working as a teacher for 5 years or more after. Low PAA scores may not find jobs as teachers, while graduates with high PAA scores may enjoy better job alternatives than working in a school.

Table 12 also shows that the number of students graduating from education majors increased about fourfold (from 3K in 1995 to 13K in 2007). And the likelihood of working in schools in years after does depends on the years after graduation. It fluctuates between 0.29 for those graduated in 1995 and 0.51 for the 2001 graduates, and then goes to 0.41 for the 2013 graduates.

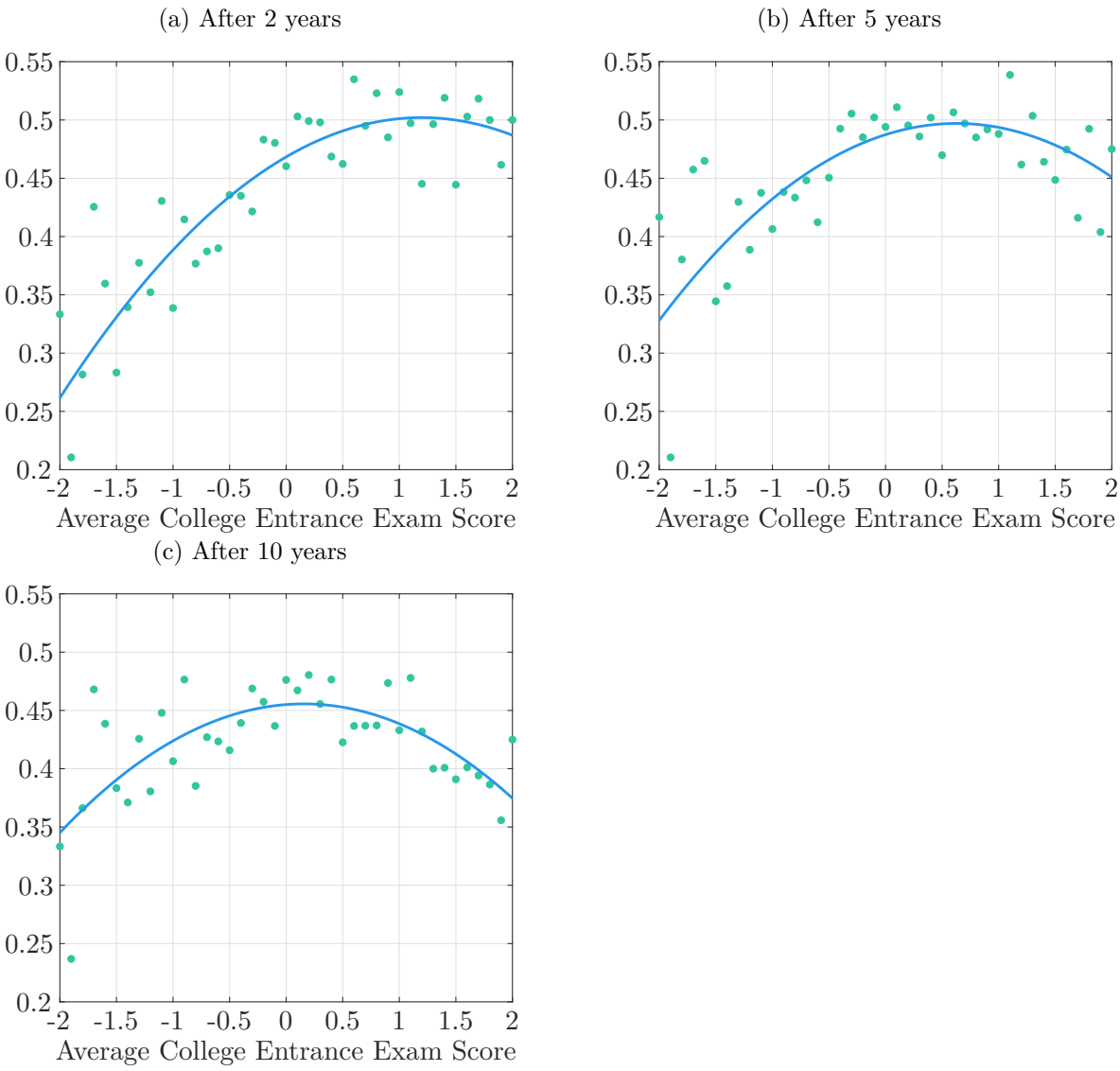
On the other hand, the graduates' PAA scores exhibit a steady tendency to decrease overtime on average. The PAA Scores for each cohort of test-takers have a mean of approx. 500 points and a standard deviation of about 100 points. There fore, the scores presented in the column can be interpreted as PAA Scores decreasing about .37 SD when comparing 1995 graduates with 2013 graduates.

Table 12: Graduates from Education working in Schools in 2011

Grad. Cohort	N	Working in 2015	PAA Score
1995	3062	0.29	539.76
1998	3707	0.35	534.66
2000	3639	0.35	529.19
2001	5051	0.33	519.53
2005	9482	0.39	506.9
2006	10068	0.42	511.65
2007	13403	0.44	507.95
2008	14528	0.45	501.07
2009	15545	0.46	495
2010	11719	0.53	504.53
2011	11836	0.54	502.65
2012	11620	0.54	514.21
2013	13390	0.51	512.52
Total	127050	0.46	508.66

Notes: This table presents information from education major graduates in years 1995, 1998, 2000, 2001 and 2005 to 2009 with valid PAA Scores (92% of the total of graduates). The PAA Scores for each cohort of test-takers has a mean of approx. 500 points and a standard deviation of about 100 points.

Figure 8: Works after graduation

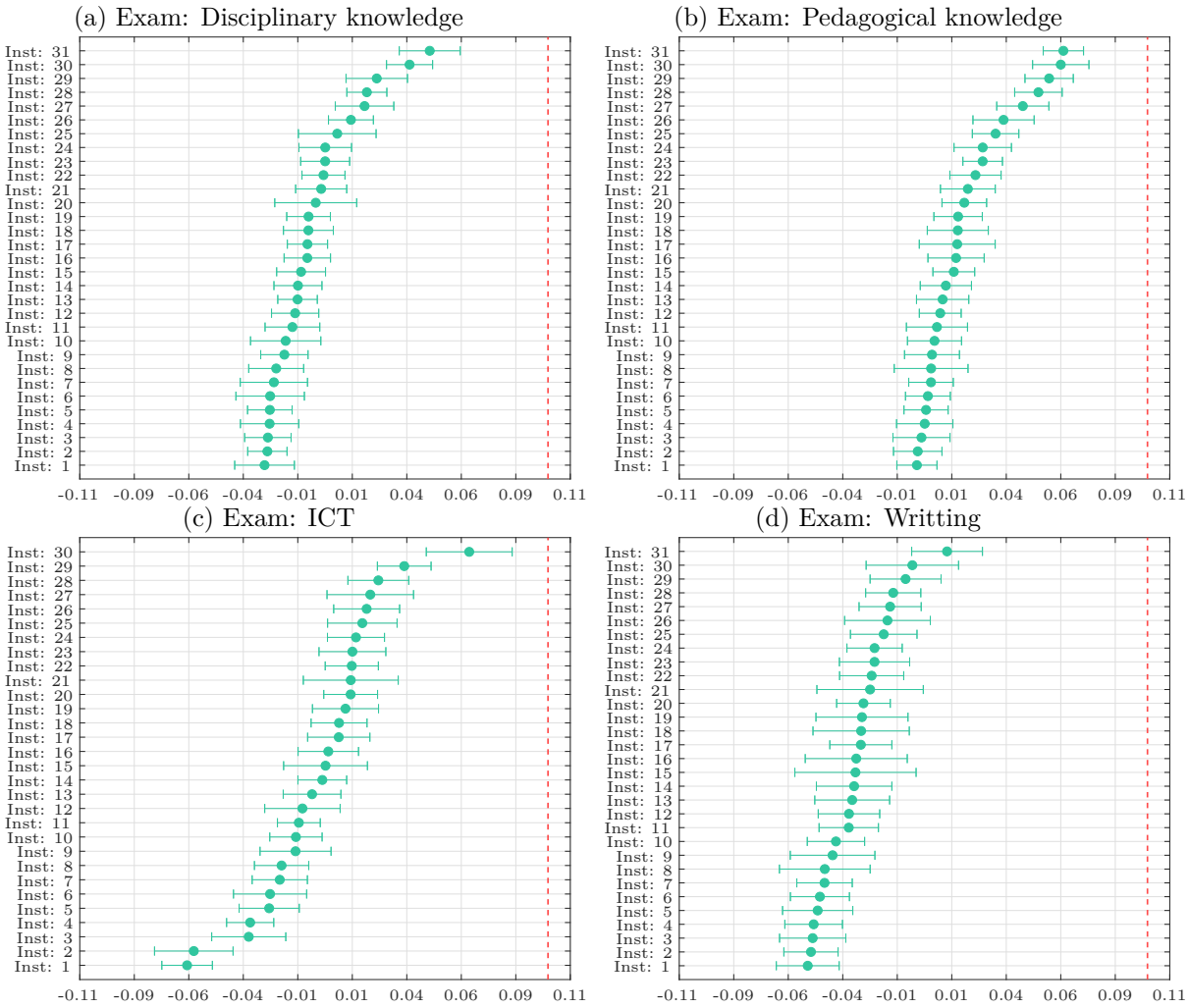


Figures 8(a), 8(b) and 8(c) plot the fraction of teachers employed in schools after 2, 5 and 10 years of graduation. The dots are averages of the outcome variable within 100 equal-sized bins of the average college entrance exam score. The data consists in graduates from teacher colleges in years 1995 to 2017, who are employed (or not) between 2003 to 2018. In the Figures the sample size is $N = 240,549$

E College Value Added

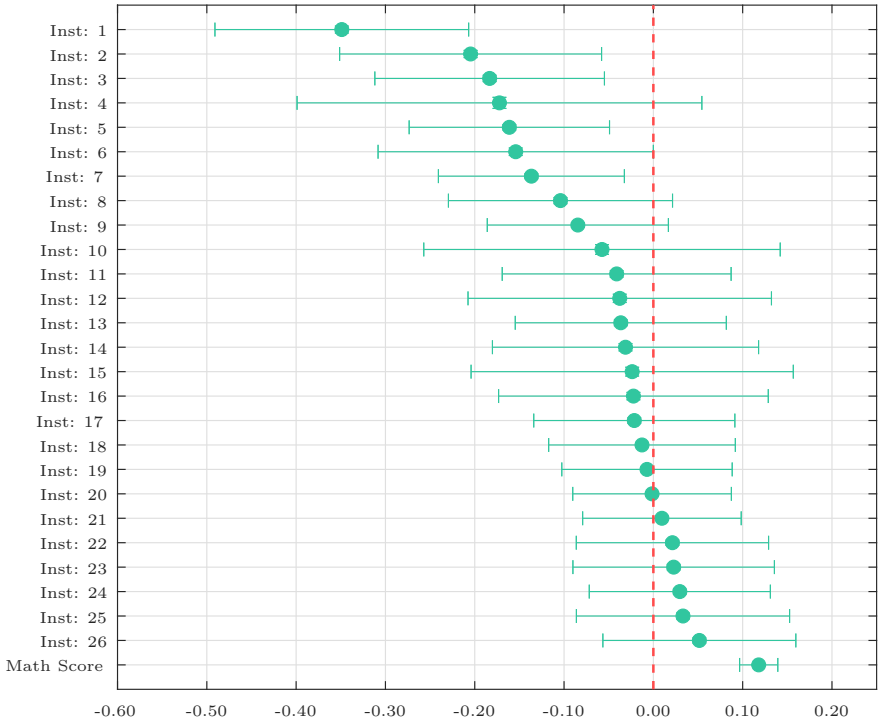
A straightforward way to assess the added value of higher education institutions is by examining more closely their estimated coefficients, which we plot in Figure 9. After controlling for PSU score, at the 5% significance level only three institutions add value to the Disciplinary exit exam score (Figure 9(a)), but the vast majority appears to add no value, in that test or the others. No institution's coefficient is statistically different from zero at the 1% significance level.

Figure 9: Institution Fixed Effects



Note: Each plot shows top 30 institution fixed effects (FE) with at least 100 observations in our sample, with 95% confidence intervals. Coefficients are sorted in descending order.

Figure 10: Program Value Added over Portfolio Scores - RD coefficients



Note: The plot shows top 26 institution treshold crossing effects for students that got accepted into each education program versus the rest of programs of education. Standard errors are computed at the 95% confidence.

F Government Policy: Recruiting

F.1 Description of the program

The *Beca de VocaciÃşn Profesor* (BVP) objective is to attract high-scoring applicants to enroll at teacher colleges. The scholarship is offered in two ways:

- Type 1: Offered to students that enter a pedagogical career as freshman students with PSU exam taken the previous year.
- Type 2: Offered to senior college students that are looking to follow a program of pedagogical formation.

F.2 Benefits of BVP

Scholarship type 1 - pedagogy

- A scholarship to cover enrollment tuition fees for students that scored over 600 on average between the mathematics and verbal exam in PSU.
- A scholarship to cover enrollment tuition fees plus 80 000 pesos (or 100 dollars) each month for students that scored over 700 on average between the mathematics and verbal exam in PSU.
- A scholarship to cover enrollment tuition fees plus 80 000 pesos (or 100 dollars) each month and a semester in a education program in a foreign country for students that scored over 600 on average between the mathematics and verbal exam in PSU.

Scholarship type 2 - degree

- Offer a scholarship for the tuition fee of the last year of the program and the pedagogical year for students that look for a pedagogical career and obtained more than 600 points in PSU.
  - For students with PSU scores higher than 600 points, a scholarship for tuition and enrollment.
  - For students with PSU scores higher than 700 points, a scholarship for tuition and enrollment and 80 000 pesos (or 100 dollars) monthly.
- Bachelors degrees other than pedagogy do not receive BVP financial aid.

F.3 BVP Requirements

- No socioeconomic requirements.
- Chilean citizenship.
- Applied to pedagogical careers through the *Council of Chancellors of Chilean Universities* (CRUCH) admission system or be admitted into a regular pedagogical career in accredited institutions or in the process of accreditation according to the *ComisiÃşn Nacional de AcreditaciÃşn* (CNA).

- Enter an eligible major as freshmen in the academic year 2011, with PSU taken in December 2010, independent from the graduation year from high school. Students who were enrolled in teacher colleges previously are not eligible.
- Obtain at least 600 points in PSU (Weighted average of 50% mathematics and 50% language) or obtain the Academic Excellence Scholarship (BEA) and a PSU score over 580.
- Enrollment in an institution and accredited career for at least two years and with cutoff score higher than 500 points on PSU as declared by the university.

F.4 Institution Requirements

- The career has to be accredited for at least two years.
- The admission cutoff score declared by the career for the *Oferta AcadÃľmica 2011* has to be of 500 points at least (50% Language and 50% Mathematics),
- Only a 15% of students can be accepted by special admission according to previously defined conditions by the Ministry of Education (including students with supernumerary vacant or Academic Excellence Scholarship).
- Only regular pedagogical careers are eligible for this program. Distance and other special programs are not eligible for thes BVP.

F.5 Procedures

There are four stages to consider: application, pre-selection of beneficiaries, selection of beneficiaries, and appealing process.

1. Application. During October, applicants must complete a form at the website www.becavocaciondeprofesor.cl or www.fuas.cl. They provide academic, personal and socioeconomic information.

2. Pre-selection. The information provided by the applicants is validated with administrative data from the Ministry of Education and all applicants that comply with the assignment requirements (scores higher than 600 points, etc) enter a list of pre-selected beneficiaries. This list is published in the website www.becavocaciondeprofesor.cl. In addition, the same website publishes which institutions/majors are eligible for the benefit.

3. Selection. Students can check whether they are in the pre-selected list. Once they know their college entrance exam scores, they decide whether to enroll at eligible institutions/majors. If they do, then they enter the list of selected beneficiaries. Institutions need to send by May 31 their list of enrolled students, which the MINEDUC uses to start the payment process to the respective institution/major.

4. Appealing. Since 2012 there is an appealing process for applicants who did not make it to the selection list. They must upload supporting documents to the website www.becasycreditos.cl, and follow the respective instructions.

G Government Policy: Screening

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

G.1 Admission to Teacher Colleges

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

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

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

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

- For the admission process in years 2017 to 2022, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 50th percentile of the distribution (500 points in the average score between mathematics and language).

  - Achieved High School GPA in the top 30% of performance.

- For the admission process in years 2023 to 2025, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 60th percentile of the distribution (525 points in the average score between mathematics and language).
  - Achieved High School GPA in the top 20% of performance.
  - Achieved a PSU score that is at least as high as the 50th percentile of the distribution (500 points) **and** a High School GPA in the top 40% of performance.

- For the admission process in year 2026 and onwards, applicants to teacher colleges have to satisfy **at least one** of the following requirements:
  - Achieved a PSU score that is at least as high as the 70th percentile of the distribution (550 points in the average score between mathematics and language).
  - Achieved High School GPA in the top 10% of performance.
  - Achieved a PSU score that is at least as high as the 50th percentile (500 points) of the distribution **and** a High School GPA in the top 30% of performance.

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

H Machine Learning Rule

Table 13: Feature Contribution to Model Accuracy (AUC)

	Graduation	Works 7 years	Works in high VA
Model 1: Only PSU scores (Training sample: 65491)			
Logistic Regression	54.17%	64.53%	65.06%
Random Forest	57.33%	64.93%	65.25%
Bagging Regressors	58.65%	60.77%	60.33%
Adaboost Classifier	59.05%	65.02%	65.24%
Gradient Boosting	65.11%	65.14%	65.35%
Model 2: PSU scores and Transcripts (Training sample: 52716)			
Logistic Regression	53.99%	65.51%	65.96%
Random Forest	56.64%	65.69%	65.91%
Bagging Regressors	57.46%	63.68%	63.20%
Adaboost Classifier	58.66%	65.84%	65.90%
Gradient Boosting	64.41%	66.05%	66.17%
Model 1: PSU scores, Transcripts and SES (Training sample: 24778)			
Logistic Regression	60.91%	65.38%	66.23%
Random Forest	63.54%	65.19%	65.80%
Bagging Regressors	64.17%	63.80%	63.44%
Adaboost Classifier	62.57%	64.89%	65.31%
Gradient Boosting	64.66%	65.51%	66.13%

Note: The table shows the area under the curve estimated for different machine learning algorithms (Logistic Regression, Random Forest, Bagging Regressors, Adaboost Classifier, and Gradient Boosting) over three different versions (PSU, PSU + Transcripts, PSU + Transcritps + SES) and using different outcome variables (Graduation, Working in Schools after 7 years and Working in High Value added schools). Our estimates are results of a Grid Search over a high dimensional grid of hyperparameters whose combination was crossvalidated using 6 different sub samples from the training sample.

Table 14: Feature Contribution to Model Accuracy (AUC)

	Sample 3	Sample 2	Sample 1
High Value Added School			
PSU	64.47%	65.43%	65.69%
T+PSU	64.40%	66.11%	
NSE+T+PSU	65.45%		
Works after 7 years			
PSU	63.56%	65.78%	65.67%
T+PSU	64.37%	66.30%	
NSE+T+PSU	64.86%		
Graduates after 6 years			
PSU	60.73%	59.22%	59.51%
T+PSU	58.18%	57.59%	
NSE+T+PSU	63.94%		
Nobs Train	24778	52716	65491
Nobs Test	2754	5858	11558

Note: The table shows the area under the curve for the Gradien Boosting Machine estimated for three different models (PSU, PSU + Transcripts, PSU + Transcritps + SES) using different outcome variables (Graduation, Working in Schools after 7 years and Working in High Value added schools). Our estimates are results of a Grid Search over a high dimensional grid of hyperparameters whose combination was crossvalidated using 6 different samples.

Table 15: Machine learning contribution to screening performance

Graduation			Works after 6 years			High Value Added School		
P. 2017	P. 2020	P. 2023	P. 2017	P. 2020	P. 2023	P. 2017	P. 2020	P. 2023
Model: PSU - Sample: 65491 - Outcome: Graduation								
16.82%	30.85%	33.87%	2.51%	5.59%	7.38%	1.95%	5.40%	5.99%
Model: PSU - Sample: 65491 - Outcome: Works 6Y								
6.67%	8.05%	6.12%	11.25%	20.89%	26.04%	13.11%	24.65%	30.12%
Model: PSU - Sample: 65491 - Outcome: High VA								
7.12%	8.09%	4.62%	11.51%	21.33%	25.58%	13.97%	25.32%	30.47%
Model: PSU & Transcripts - Sample: 52716 - Outcome: Graduation								
15.62%	29.34%	33.33%	2.37%	4.87%	7.63%	-0.93%	-0.70%	4.75%
Model: PSU & Transcripts - Sample: 52716 - Outcome: Works 6Y								
6.37%	5.06%	1.59%	12.61%	22.87%	28.50%	11.67%	21.46%	30.65%
Model: PSU & Transcripts - Sample: 52716 - Outcome: High VA								
6.51%	5.17%	0.61%	12.56%	21.41%	27.74%	11.83%	21.00%	30.50%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: Graduation								
12.30%	22.73%	24.16%	-0.42%	5.03%	7.57%	-3.65%	2.52%	0.00%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: Works 6Y								
6.51%	7.27%	1.89%	9.70%	17.26%	25.66%	9.29%	17.23%	23.51%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: High VA								
6.22%	6.73%	3.99%	7.59%	17.66%	25.99%	7.79%	19.33%	24.50%

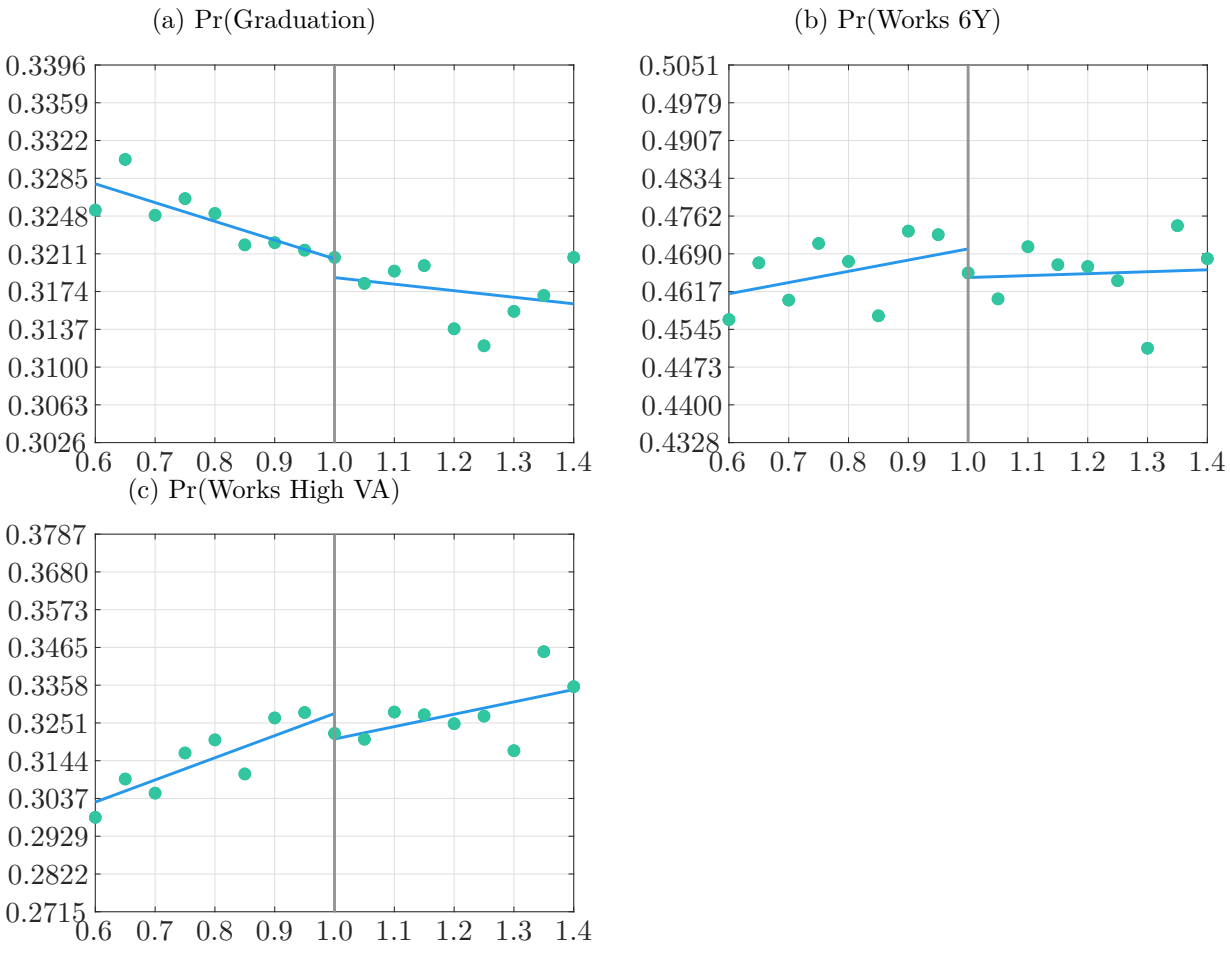
Note: The table computes the contribution of the Machine Learning rule to the performance of the screening policy with different samples. All of the models use a Gradient Boosting Machine algorithm. We test nine combinations of the model where we use different three diferent versions of input features (PSU, PSU + Transcripts, PSU + Transcripts + SES) and three different outcomes (Timely graduation, Work in schools after 7 years, Working in High Value Added) to predict performance. Finally, we use the probabilities estimated in the test sample and compare to the performance of the government policy.

Table 16: Machine learning contribution to screening performance holding sample constant

Graduation			Works after 6 years			High Value Added School		
P. 2017	P. 2020	P. 2023	P. 2017	P. 2020	P. 2023	P. 2017	P. 2020	P. 2023
Model: PSU - Sample: 24778 - Outcome: Graduation								
13.17%	26.18%	28.36%	1.27%	6.79%	9.21%	-1.66%	2.52%	1.98%
Model: PSU - Sample: 24778 - Outcome: Works 6Y								
5.93%	5.82%	1.60%	9.07%	16.71%	23.58%	8.96%	18.49%	22.13%
Model: PSU - Sample: 24778 - Outcome: High VA								
6.66%	5.82%	3.15%	8.65%	15.35%	21.55%	8.79%	16.18%	22.77%
Model: PSU & Transcripts - Sample: 24778 - Outcome: Graduation								
11.00%	20.91%	24.58%	0.42%	4.62%	5.59%	-2.32%	-0.84%	-1.73%
Model: PSU & Transcripts - Sample: 24778 - Outcome: Works 6Y								
6.51%	4.73%	0.21%	9.28%	16.85%	23.85%	8.46%	17.86%	22.52%
Model: PSU & Transcripts - Sample: 24778 - Outcome: High VA								
5.07%	6.73%	1.47%	8.54%	16.44%	23.03%	8.29%	17.44%	24.01%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: Graduation								
13.46%	23.27%	27.10%	-1.79%	4.76%	8.22%	-5.31%	2.10%	1.98%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: Works 6Y								
6.51%	7.27%	1.89%	9.70%	17.26%	25.66%	9.29%	17.23%	23.51%
Model: PSU & Transcripts & SES - Sample: 24778 - Outcome: High VA								
6.22%	6.73%	3.99%	7.59%	17.66%	25.99%	7.79%	19.33%	24.50%

Note: The table computes the contribution of the Machine Learning rule to the performance of the screening policy holding the sample constant at 24778 observations. All of the models use a Gradient Boosting Machine algorithm. We test nine combinations of the model where we use different three diferent versions of input features (PSU, PSU + Transcripts, PSU + Transcripts + SES) and three different outcomes (Timely graduation, Work in schools after 7 years, Working in High Value Added) to predict performance. Finally, we use the probabilities estimated in the test sample and compare to the performance of the government policy.

Figure 11: Discontinuity over probabilities estimated using ML



Note: The figures show the treshold crossing effect over probabilities estimated using a Gradiend Boosting Machine whose input features are Entry Exam scores and students' transcripts. The estimates of the figure correspond to students from 2011 sample.