Screening and Recruiting Talent At Teacher Colleges

Franco Calle, Sebastian Gallegos and Christopher A. Neilson

NBER Education
Cambridge, MA
November 14th, 2019

Coauthors

Franco Calle Falcón
Senior Research Specialist
Princeton University
Affiliated Researcher at ConsiliumBots.
Industrial Organization, Labor and Data Science.
Site: https://francocalle.github.io/ @FrancoCalle

Coauthors

Franco Calle Falcón
Senior Research Specialist
Princeton University
Affiliated Researcher at ConsiliumBots.
Industrial Organization, Labor and Data Science.
Site: https://francocalle.github.io/ @FrancoCalle

Sebastián Gallegos
Research Economist
Inter-American Development Bank
Ph.D., University of Chicago Harris School
Development, Education, and Behavioral Economics.
Site: scholar.princeton.edu/sebagallegos @seba_gallegos

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.
   - **Retaining** Talented Teachers in the teaching career path.

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.
   - **Retaining** Talented Teachers in the teaching career path.
   - **Training** at Teacher Colleges and Developing Skills On the Job

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.
   - **Retaining** Talented Teachers in the teaching career path.
   - **Training** at Teacher Colleges and Developing Skills On the Job
   - **Increasing Effort** through Incentives like Pay-For-Performance

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.
   - **Retaining** Talented Teachers in the teaching career path.
   - **Training** at Teacher Colleges and Developing Skills On the Job
   - **Increasing Effort** through Incentives like Pay-For-Performance

Each of these policies have important practical challenges, especially in developing countries.

Motivation - Teacher quality matters, policies focus on it

1. Teacher quality matters (Rockoff 2004, Aaronson et al. 2007, Chetty et al. 2014, Araujo, et al 2018)
2. Ways to improve the systems teacher productivity (Jackson 2012, Jackson et al (2014))
   - **Recruiting** Talented students into the teaching profession.
   - **Retaining** Talented Teachers in the teaching career path.
   - **Training** at Teacher Colleges and Developing Skills On the Job
   - **Increasing Effort** through Incentives like Pay-For-Performance

Each of these policies have important practical challenges, especially in developing countries.

Proposition: Increased availability of data and decreasing cost of prediction suggests
$\Rightarrow$ **Recruiting/Screening** policies might increasingly become more feasible and effective going forward.

This paper - What we do

1. Use large administrative datasets to document that pre-college academic achievement is significantly correlated with measures of teacher quality

   ⇒ **Concave relationship** between pre-college academics and various measures of productivity.

This paper - What we do

1. Use large administrative datasets to document that pre-college academic achievement is significantly correlated with measures of teacher quality

   ⇒ **Concave relationship** between pre-college academics and various measures of productivity.

2. Evaluate the medium run effectiveness of a recruiting policy targeted at high performing high school students.

   ⇒ RD design - **big and persistent effects** on individual choices, modest effects in aggregate.

This paper - What we do

1. Use large administrative datasets to document that pre-college academic achievement is significantly correlated with measures of teacher quality

   ⇒ **Concave relationship** between pre-college academics and various measures of productivity.

2. Evaluate the medium run effectiveness of a recruiting policy targeted at high performing high school students.

   ⇒ RD design - **big and persistent effects** on individual choices, modest effects in aggregate.

3. Evaluate short run impact and simulate the long run effects of a recent screening policy.

   ⇒ Simulation of screening policy based on prediction model validated by the RD, **larger aggregate effects**.

This paper - What we do

1. Use large administrative datasets to document that pre-college academic achievement is significantly correlated with measures of teacher quality

   ⇒ **Concave relationship** between pre-college academics and various measures of productivity.

2. Evaluate the medium run effectiveness of a recruiting policy targeted at high performing high school students.

   ⇒ RD design - **big and persistent effects** on individual choices, modest effects in aggregate.

3. Evaluate short run impact and simulate the long run effects of a recent screening policy.

   ⇒ Simulation of screening policy based on prediction model validated by the RD, **larger aggregate effects**.

4. Show how governments can use administrative data can help drive design details for **screening and recruiting policies**.

This paper - How we do it

Historical Data from Chile: The population of college entrance exams (1967-), teacher's performance measures, and earnings.

1. Big data $\Rightarrow$ nonparametric plots to explore correlations
2. Policy allows for RD design $\Rightarrow$ to assess recruiting policies
3. Data allows for medium run outcomes for RD design $\Rightarrow$ Assess invariance to policy of correlations
4. Machine Learning $\Rightarrow$ Prediction to simulate effectiveness of recent screening policies

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.
- Low student teacher-ratio (approx 20, close to OECD levels)

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.
- Low student teacher-ratio (approx 20, close to OECD levels)
- Low teacher absenteeism (5% vs 20% in other developing countries)

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.
- Low student teacher-ratio (approx 20, close to OECD levels)
- Low teacher absenteeism (5% vs 20% in other developing countries)
- $\Rightarrow$ Enough teachers in classrooms, but

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.
- Low student teacher-ratio (approx 20, close to OECD levels)
- Low teacher absenteeism (5% vs 20% in other developing countries)
- $\Rightarrow$ Enough teachers in classrooms, but
  1. Paid less than other professions

Context : Teacher stock in Chile

- The teacher labor force in Chile (2018) is composed by approx 200K classroom teachers, for around 3 million students in grades 1 to 12.
- The number of teachers has increased sharply while the population of students has remained stable.
- Low student teacher-ratio (approx 20, close to OECD levels)
- Low teacher absenteeism (5% vs 20% in other developing countries)
- $\Rightarrow$ Enough teachers in classrooms, but
  1. Paid less than other professions
  2. Students with higher college entrance scores are less likely to enter teaching

Data Sources

Pre-college academic achivement

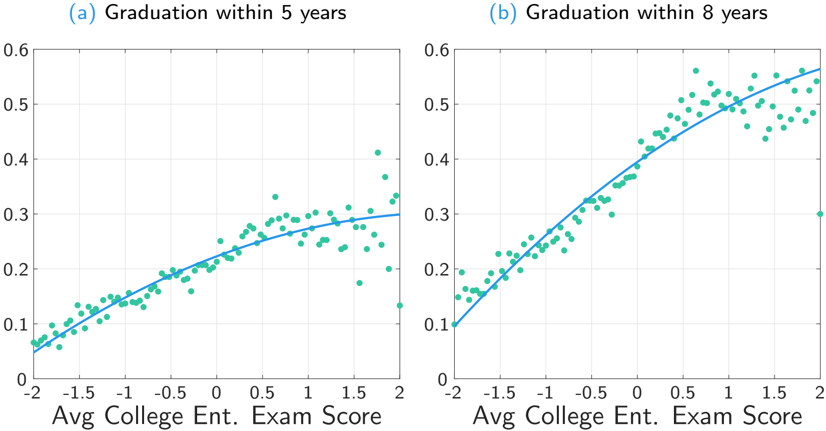
- High School Course Transcripts and GPA.
- College entrance exams (many subjects).

Teacher Outcome Measures

- Young Teachers: **Exit exam** (INICIA test).
- Public/Private Employment roster: **Employment** (Idionidad docente)
- Teacher Productivity: **Teacher Evaluation, Value Added** (Evaluacion Docente, ED, SEPA DeGregorio, Gallegos and Neilson (2019))
- Teacher Productivity: **Wages** (SII Earnings)

- Other measures such as School Value Added, School Characteristics, Peer Teacher Characteristics, Student Test Scores - Neilson (2014)

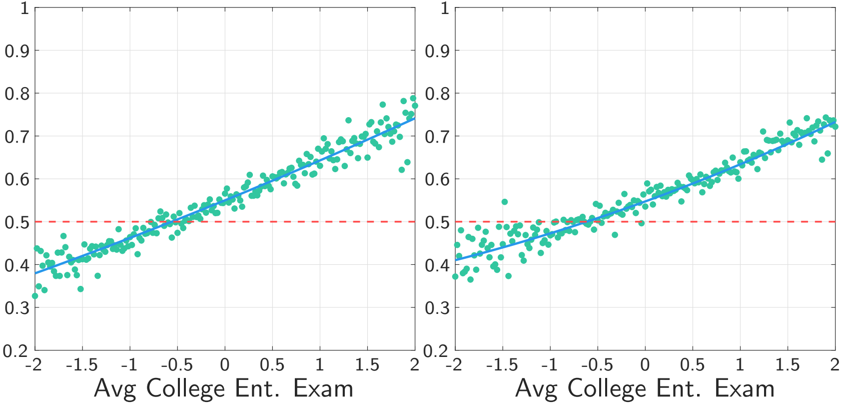
College Entrance Exam Score And **Graduation**



Note: The figures plot probability of graduation for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data.

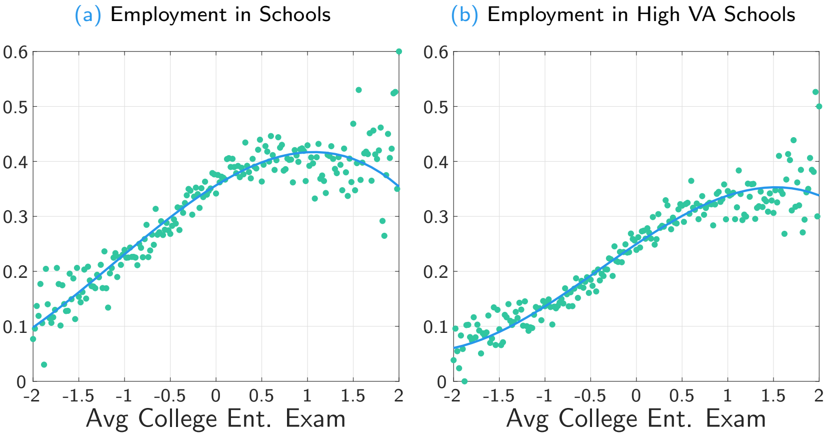
College Entrance Exam Score And **College Exit Exams**

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



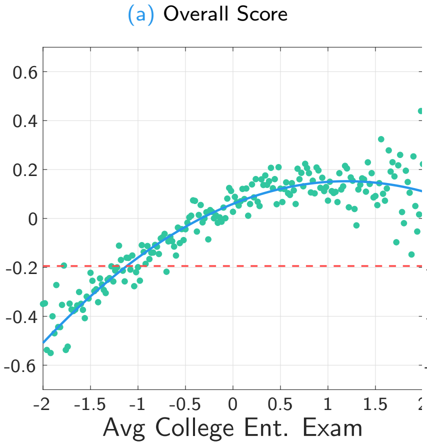
Note: The figures plot the mean of each exit exam for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data. The data consists in graduates who took the respective exit exam test between years 2009 and 2017.

College Entrance Exam Score And **Working in Schools**



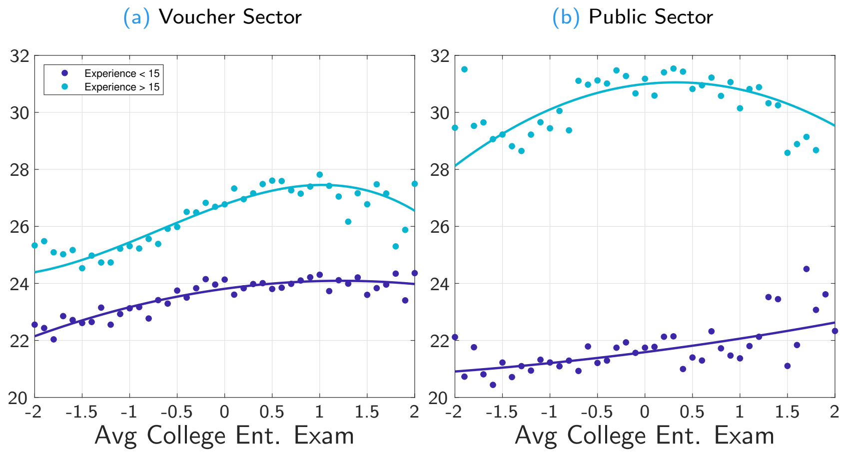
Notes: The figures plot the mean of the y-axis variable within 100 equal-sized bins of the average college entrance exam score, and fit estimated lines using all the underlying data.

College Entrance Exam Score And **In-Class Evaluation**



Note: The figures plot the scores in the teacher's evaluation for 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data.

College Entrance Exam Score And **Wages**



Note: The figures plot the wages for teachers in the public and private sector in dollars in 100 equal-sized bins of the average college entrance exam score fits an estimated line using the underlying data.

Summary table: Teaching Performance vs PSU Scores

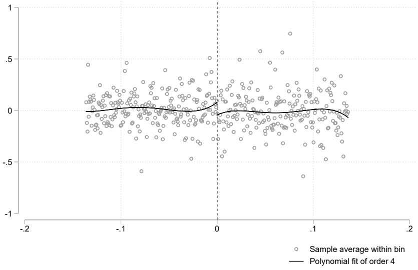
Table: Summary table: Teaching Performance vs PSU Scores

Graduation	Years after enrollment			
	5 Years	8 Years		
PSU Score	0.073***	0.118***		
	(0.002)	(0.002)		
(PSU Score)^2	-0.027***	-0.026***		
	(0.001)	(0.001)		
Observations	[84,847]	[84,847]		
Dep. Var. Mean	0.322	0.473		
Exit Exams	Disciplinary Test	Pedagogy Test	Writing Test	ICT Test
PSU Score	0.509***	0.506***	0.463***	1.27 ***
	(0.005)	(0.007)	(0.007)	(0.014)
(PSU Score)^2	0.043***	0.033***	-0.021***	-0.07 ***
	(0.003)	(0.311)	(0.200)	(0.443)
Observations	[35,355]	[33,409]	[11,300]	[5,517]
Dep. Var. Mean	0.000	0.000	0.000	
Productivity Measures:	Teacher Evaluation Overall	Teacher Evaluation Portfolio	Wages in Public Schools	Wages in Private Schools
PSU Score	0.615 ***	0.477 ***	0.536 ***	0.628 ***
	(0.041)	(0.04)	(0.046)	(0.043)
(PSU Score)^2	-0.048 ***	-0.031 ***	-0.049 ***	-0.055 ***
S	(0.001)	(0.001)	(0.002)	(0.002)
Observations	[63539]	[63539]	[36771]	[58523]
Dep. Var. Mean	0.000	0.000	0.000	0.000
Employment		Years after graduation		Value

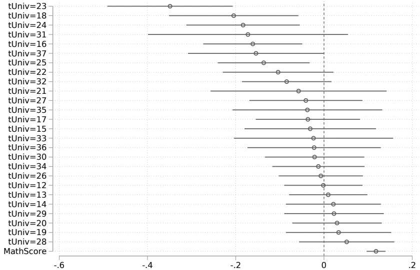
Summary table: Teaching Performance vs PSU Scores

- The correlation between entrance exams and outcomes could be due to unequal access to high value added colleges.
- We use the centralized college assignment mechanism to generate an RD desig to study college VA.
- We find there is no evidence that universities add more or less value.

Education Institutions' Value Added to Teacher Evaluation



(a) Local Polynomial RD estimation on college teaching majors
(X: Distance to the cutoff; Y: Teacher evaluation score)



(b) Institutions' RD Threshold Crossing Effects

Taking Stock

1. Low achievement students seem to systematically be associated with low performance in teaching.
2. Relationship seems concave.
3. Colleges do not seem to be generating the differences.

$\Rightarrow$ Can we move towards screening and recruiting policies using this correlation?

$\Rightarrow$ Are these correlations persistent and policy invariant?

A Recruiting and Screening Policy: *Beca Vocacion Profesor* (BVP)

- This policy gave full scholarships and other incentives such as stipends and paid semesters abroad for students who matriculate at teaching colleges with scores from approximately the highest 30% of the admissions test distribution.
- Also teacher colleges needed to implement a cutoff score of the 50th percentile of the average score distribution.

Carrot and Stick Incentives

For students:

Carrot and Stick Incentives

For students:

- 600 points (top 25%) - full ride scholarship.

Carrot and Stick Incentives

For students:

- 600 points (top 25%) - full ride scholarship.
- 700 points (top 5%) - full ride + **monthly stipend** of approx $US150 (approx 50% minimum wage).

Carrot and Stick Incentives

For students:

- 600 points (top 25%) - full ride scholarship.
- 700 points (top 5%) - full ride + **monthly stipend** of approx \$US150 (approx 50% minimum wage).
- 720 points (top 2%) - full ride + monthly stipend of approx \$US150 (approx 50% minimum wage) + **paid semester abroad**.

Carrot and Stick Incentives

For students:

- 600 points (top 25%) - full ride scholarship.
- 700 points (top 5%) - full ride + **monthly stipend** of approx $US150 (approx 50% minimum wage).
- 720 points (top 2%) - full ride + monthly stipend of approx $US150 (approx 50% minimum wage) + **paid semester abroad**.

For colleges

Carrot and Stick Incentives

For students:

- 600 points (top 25%) - full ride scholarship.
- 700 points (top 5%) - full ride + **monthly stipend** of approx $US150 (approx 50% minimum wage).
- 720 points (top 2%) - full ride + monthly stipend of approx $US150 (approx 50% minimum wage) + **paid semester abroad**.

For colleges

- Be accredited for at least 2 years at all campuses

Carrot and Stick Incentives

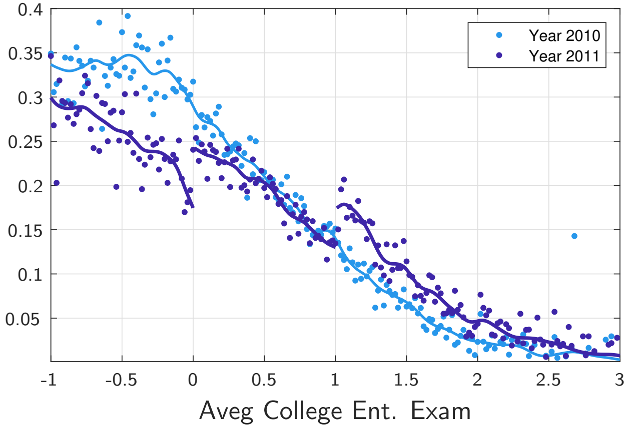
For students:

- 600 points (top 25%) - full ride scholarship.
- 700 points (top 5%) - full ride + **monthly stipend** of approx $US150 (approx 50% minimum wage).
- 720 points (top 2%) - full ride + monthly stipend of approx $US150 (approx 50% minimum wage) + **paid semester abroad**.

For colleges

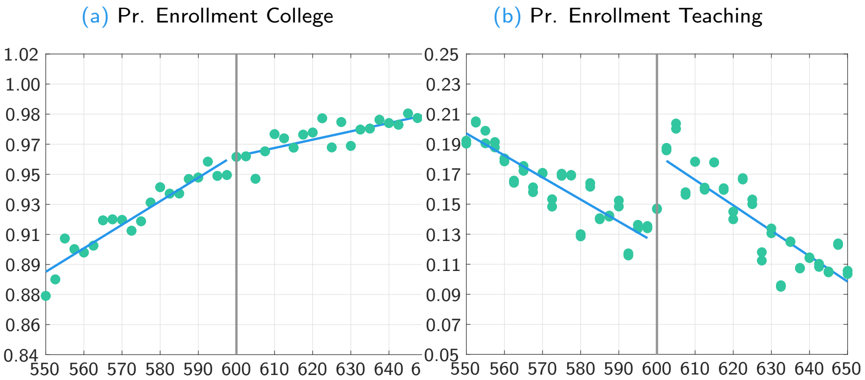
- Be accredited for at least 2 years at all campuses
- Minimum score of 500 with no more than 15% exceptions.

Individual Choice Probabilities



Note: The figure above shows the probability of enrolling in any teaching college. Color ■ represent the probabilities for 2011, while ■ the probabilities for 2010. Source: MINEDUC and DEMRE.

Threshold Crossing Effect - Choice Probability



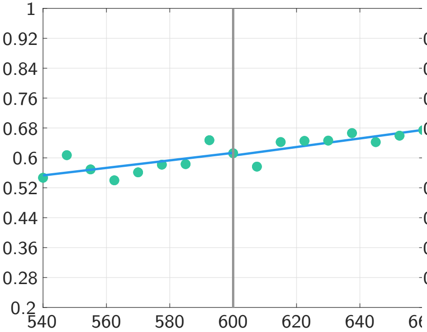
Note: Left panel plots threshold crossing effect over the number of students enrolled into teaching programs conditional on PSU score while, right panel shows probability of enrollment in any college conditional on PSU score.

RDD Regression across different BVP Thresholds

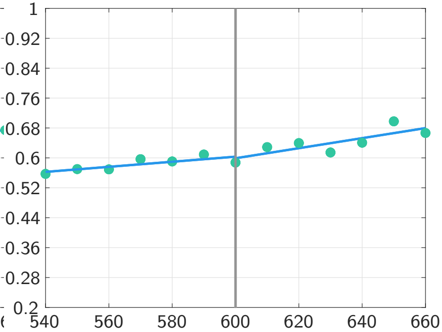
	T = 600		T = 700		T = 720	
	Coefficient	T-stat	Coefficient	T-stat	Coefficient	T-stat
Above T	0.054***	4	0.026***	2.42	-0.007	-0.72
Const.	0.120***	20.01	0.028***	4.6	0.0033***	5.38
PSU	-0.001***	-1.54	-0.0015***	-3.18	-0.005***	-1.12
N. Obs.	18007		5450		4150	

Persistent Effects : Exit Exams

(a) Pedagogy Exam



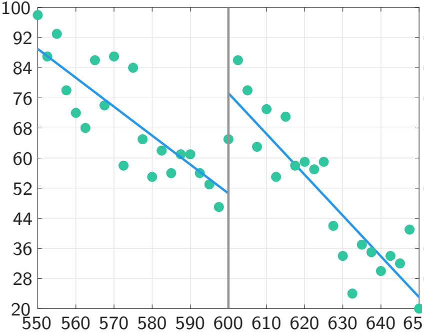
(b) Disciplinary Exam



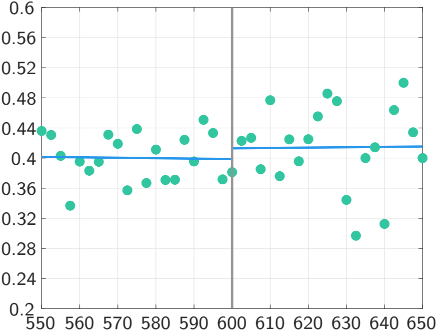
Note: The Figures above plots the threshold crossing effect over exit exams which are taken after 5 or 6 years of enrollment in 2011 conditional on psu scores.

Persistent Effects : Employment 2018, Cohort 2011

(a) Number Employed 2011

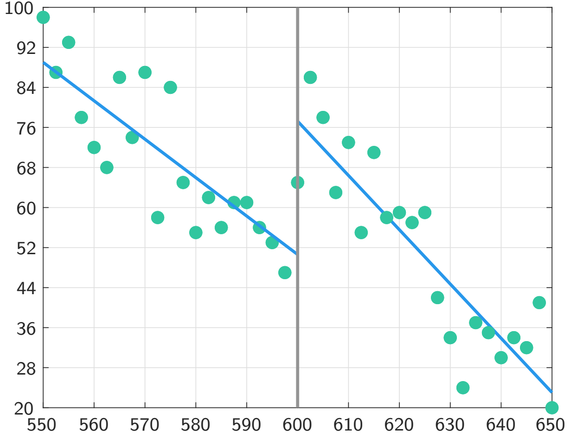


(b) Employment Rate 2011



Note: The Figures above plots the threshold crossing effect over number of teachers employed and probability of bein employed after 7 years for students enrolled in 2011 conditional on psu scores.

Number Employed Conditional on Score



Supply Side : Participation of Colleges

Supply Side : Participation of Colleges

- High take up among public institutions and traditional private colleges (CRUCH) but low or null participation from institutions with less qualified students and Professional Institutes.

Supply Side : Participation of Colleges

- High take up among public institutions and traditional private colleges (CRUCH) but low or null participation from institutions with less qualified students and Professional Institutes.
- Eligible institutions and careers covered only 40% of matriculated students in 2010.

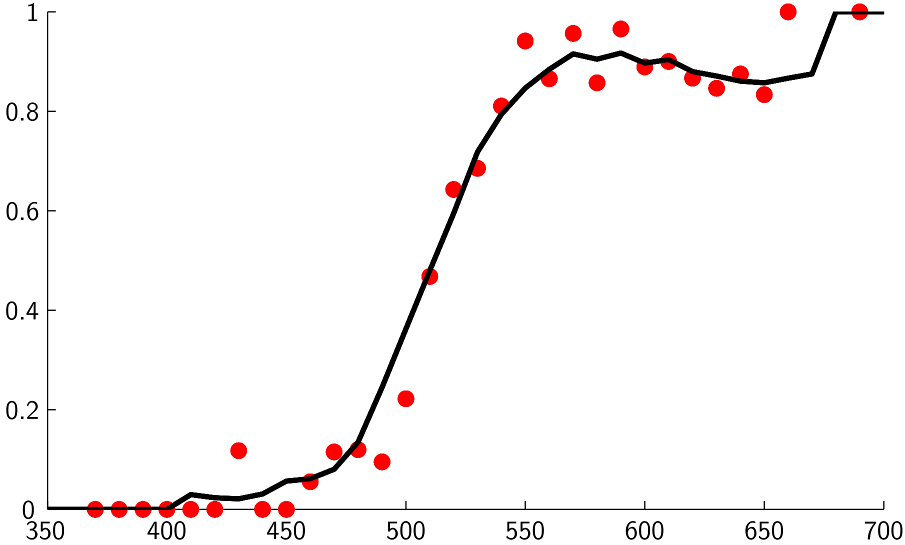
Supply Side : Participation of Colleges

- High take up among public institutions and traditional private colleges (CRUCH) but low or null participation from institutions with less qualified students and Professional Institutes.
- Eligible institutions and careers covered only 40% of matriculated students in 2010.
- Approximately 1/3 career/college combinations that were eligible did not participate.

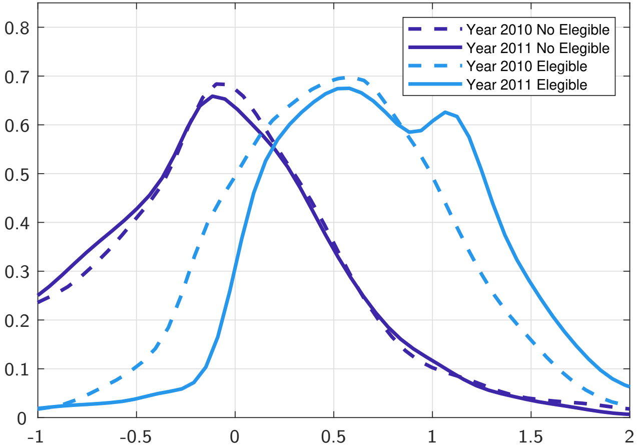
Supply Side : Participation of Colleges

- High take up among public institutions and traditional private colleges (CRUCH) but low or null participation from institutions with less qualified students and Professional Institutes.
- Eligible institutions and careers covered only 40% of matriculated students in 2010.
- Approximately 1/3 career/college combinations that were eligible did not participate.
- Approximately 1/4 career/college combinations were not even eligible.

Supply Side : Participation of Colleges

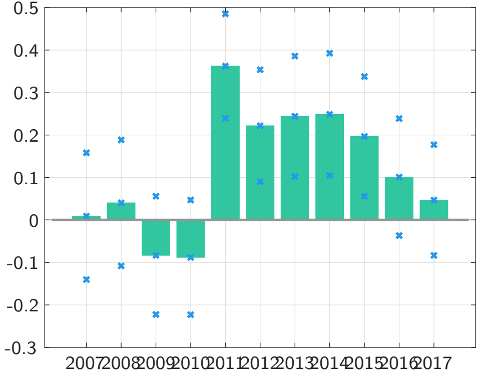


Aggregate Effects on the Distribution of Scores



Note: In the figure above the continuous and dotted lines show the scores distribution for year 2010 and 2011 respectively. The ■ represents scores distribution for non BVP schools while the ■ shows the distribution for BVP schools.

Over Time - RDD Estimates on Freshmen Enrollment in Teacher Colleges



Taking Stock

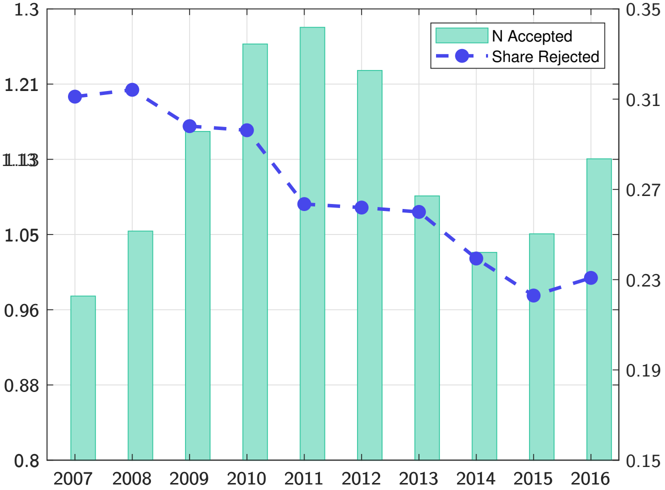
1. Recruiting Policy seemed to work to exclude low achievement students by inducing voluntary minimum requirements.
2. Recruiting policy increased probability of high achieving students by a lot but in levels this has small effect on stock.
3. The effects of the policy are more or less eliminaed when college becomes free overall.
4. Less strict policy may have had bigger impact by increasing take up by colleges.
5. $\Rightarrow$ Relationship between scores and medium run outcomes seems persisitent and invarient to policy.

Nueva Ley de Carrera Docente (NLCD) policy

- This is a broad policy implemented in 2017 created a new system of professional development for teachers in the country

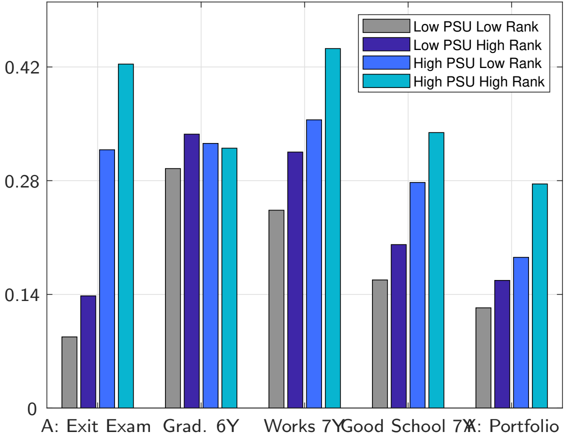
- One important component: barred all teaching colleges from admitting students with **below average scores** <u>unless</u> they had **high GPA**

Rule Applied To Past Teaching Students



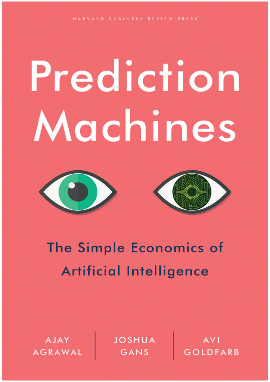
Note: In the Figure above ■ shows the share of students that would have been rejected by the policy, meanwhile ■ shows the number of students (in thousands) that would have been accepted by the rule.

Outcomes for Those Screened In Simulation



Note: The figure above shows the labor outcomes for each group of students enrolled in pedagogy from 2007 - 2016.

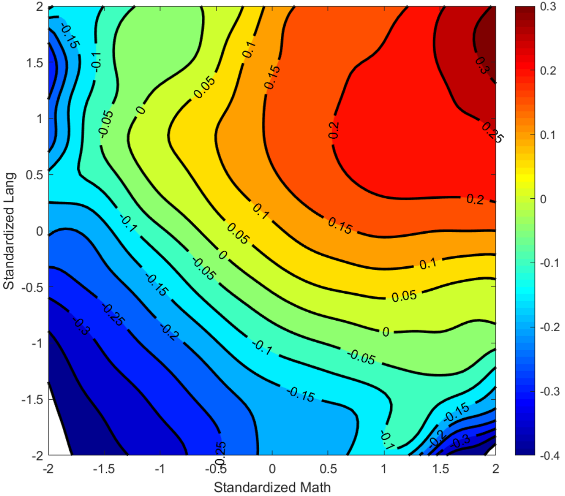
The cost of prediction has gone down



Machine Learning can be thought fundamentally as a *prediction technology*. Increased availability of data and advances in ML mean that **the cost of prediction is going down.**

Building on evidence presented, we now revisit policy questions equipped with a) big data b) flexible tools for prediction.

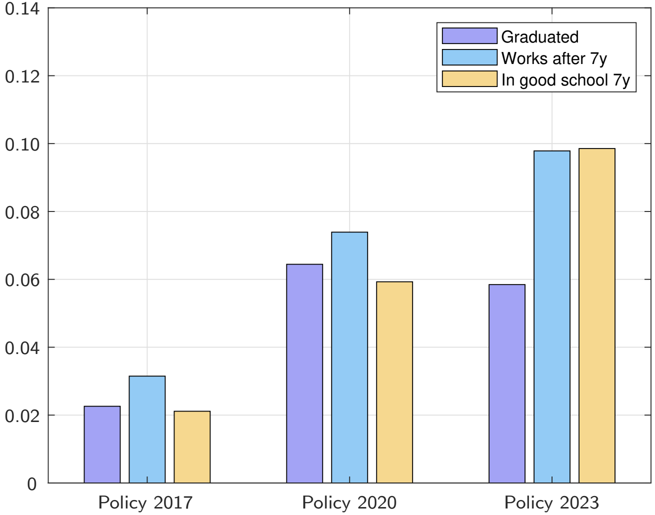
Non Parametric prediction of Bad Teachers



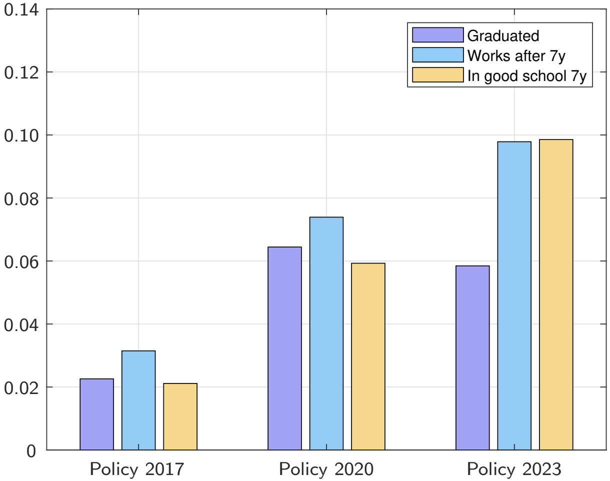
Improving Policy Leveraging More Data

- Current government rule imposes very specific and arbitrary weights on certain variables (Math, Lang, GPA).
- What about using information on the other tests or from the full student transcripts, other standardized tests?
- We estimate a series of simple policy rules : OLS, Random Forest, etc.
- We choose parameters so that the policy rules have the same Type I error as the government rule, and minimize Type II to illu

Performance Measures given same Accuracy Level as Government Rule



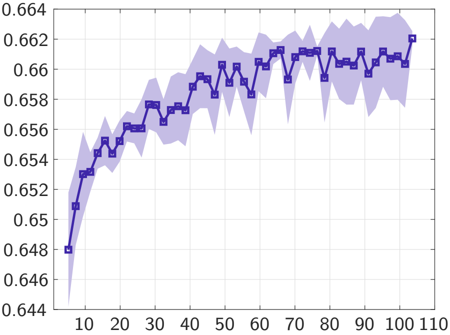
Performance given Accuracy of Government Rule



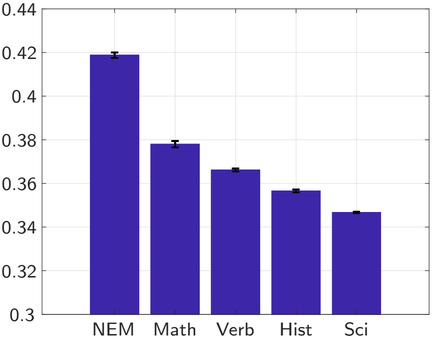
Notes: The figures above show the percentage increase in each Graduation, working after 7 years and working in a good school for the students that would have been admited by an ML screening method with a count of students rejected equivalente to those screened out by the rules proposed by the government.

Variable Contribution to Performance

(a) Data contribution



(b) Variable contribution



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

Taking Stock

1. Low achievement students seem to systematically be associated with low performance in teaching.
2. Relationship seems concave in the context of Chile.
3. From Screening and Recruiting Policies
   - Can work but most effective when excluding low-scoring students.
   - Can recruit but increasing very high ability students is hard.
4. Simple rules can potentially work, but should be Data can be helpful in guiding the screening policy.