

Perceived Ability and School Choices

Experimental Evidence and Scale-up Effects

Matteo Bobba (TSE) Veronica Frisancho (CAF) Marco Pariguana (Edinburgh)

NBER Summer Institute
Development Economics

July 2023

Motivation

- Human capital investments rely on expected **individual-specific net returns**
- Providing (aggregate) info in education markets affect schooling decisions
 - School quality (Hastings-Weinstein, 2008; Mizala-Urquiola, 2014)
 - Application procedures (Hoxby-Turner, 2014; Dinkelman-Martinez, 2014)
 - Labor market returns (Jensen, 2010; Wiswall-Zafar, 2015)
- How does information about **own academic skills** shape choices/outcomes?
⇒ **What are the effects of such information intervention at scale?**

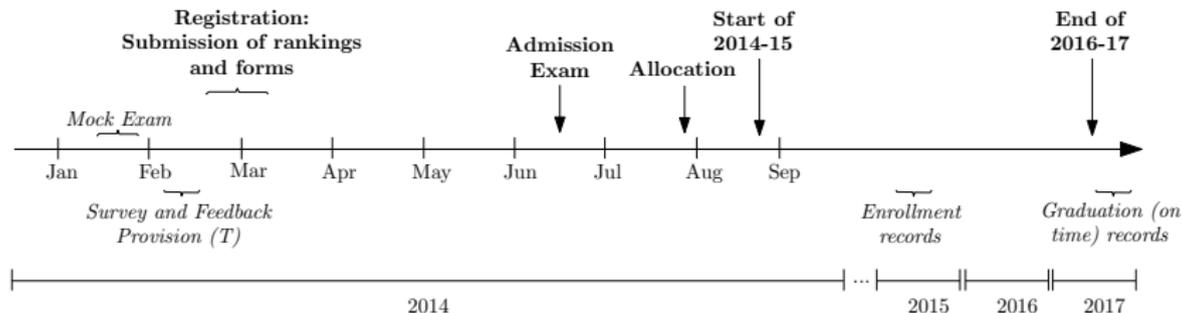
This Paper

- 1 Field experiment provides students with feedback on their academic skills
 - Quantify impacts on beliefs, school choice/placement and **educ. outcomes** (Dizon-Ross, 2019; Bergman, 2021)
- 2 Estimate a model of school choice to scale-up information intervention
 - Simulate policy impacts at scale **within a centralized assignment mechanism** (Andrabi-Das-Khwaja, 2017; Allende-Gallego-Neilson, 202x)
 - Link model-based sorting to realized discontinuities in admission probabilities (Barrow-Sartain-De la Torre, 2020; Angrist-Pathak-Zarate, 2023)

Centralized School Assignment in Mexico City

- Centralized admission system for public high schools
 - \approx 300k applicants and more than 600 schools
 - Matching mechanism is the Serial Dictatorship
 - ROLs are submitted before students know their admission exam score
- High-school tracks: General, Technical/Vocational, and Elite Cutoff Scores by Track
 - General track is oriented toward college education
 - 32 elite schools are affiliated with two prestigious higher educ. institutions

Information Experiment



- RCT sample consists of 9th graders in low-SES neighborhoods
 - $\approx 2,500$ applicants in 90 schools ($\approx 1\%$ of total) [Spatial Distr.](#) [Population vs. Sample](#)
 - Mock test scores are informative signals [More on Mock Test](#) [More on Beliefs](#)

Experimental Evidence on Performance Feedback

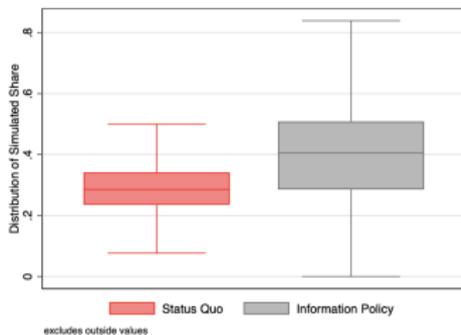
- 1 Perception gaps about test performance shrink by one third
 - Larger updates among lower performing students [Beliefs Charts](#) [Regression Results](#)
- 2 Shift in school choices → sorting by skill across high-school tracks
 - No average effect of the information intervention [Regression Results](#)
 - No effect on elite schools
- 3 7% increase in high-school graduation on time wrt control group
 - Larger effect for lower performing kids [High-School Stats](#) [Graduation Chart](#) [Regression Results](#)

Using Model+Experiment to Scale-up the Intervention

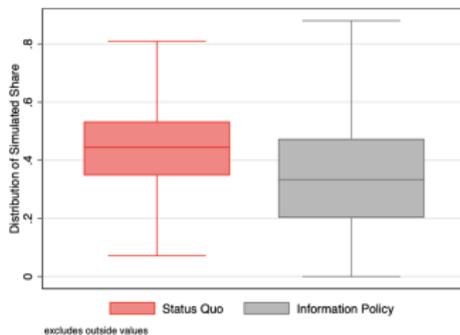
- Recover the distribution of preferences over schools in RCT and extrapolate
 - Replace the individual scores in the mock test with those in the adm. exam
 - Match = $f(\text{Simulated Prefs}, \text{Exam Scores}, \text{School Capacities})$
- Pref. parameters for the control group likely capture the status quo scenario
 ⇒ Validation using realized match Assignment Outcomes Cutoff Scores
- Pref. parameters for the treatment group approximate the scaled-up policy
 ⇒ Changes in matching eq. driven by demand-side Assignment Outcomes Cutoff Scores

Feedback Provision at Scale: Demand Effect

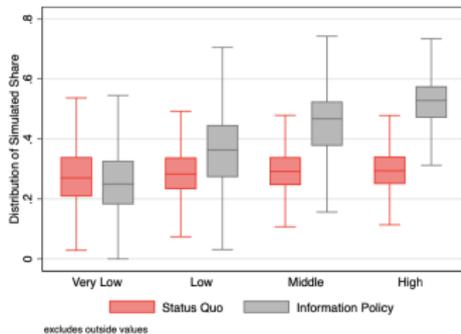
Marginal effect by poverty



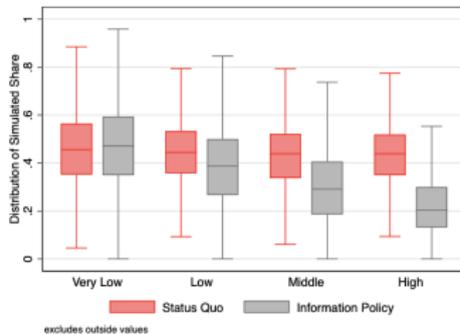
(a) Aggregate Shares of Academic Schools



(b) Aggregate Shares of Elite Schools

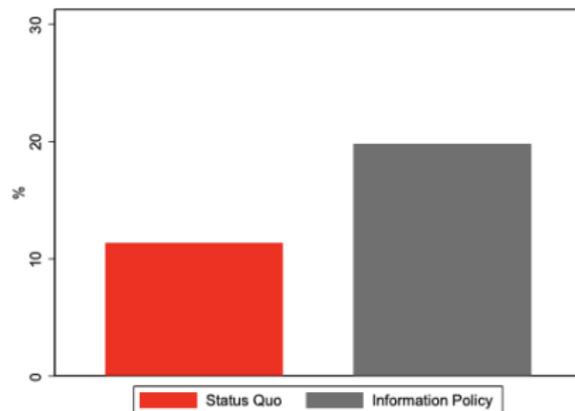


(c) Share of Academic Schools by SES

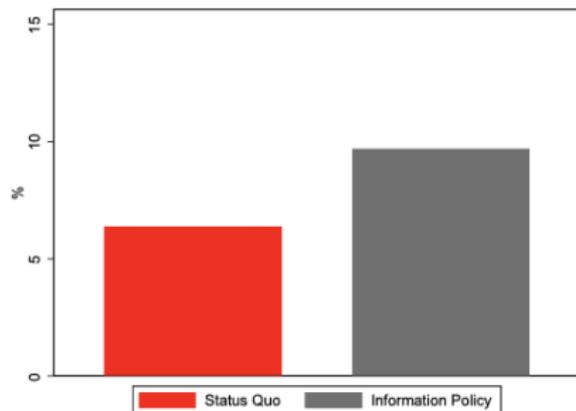


(d) Share of Elite Schools by SES

Feedback Provision at Scale: Equilibrium Effect



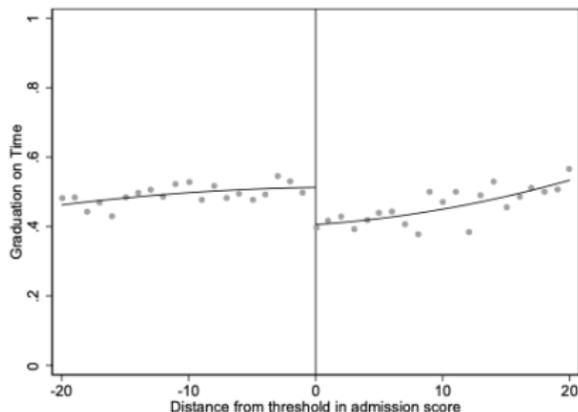
(a) % from Low-Income Neighborhoods



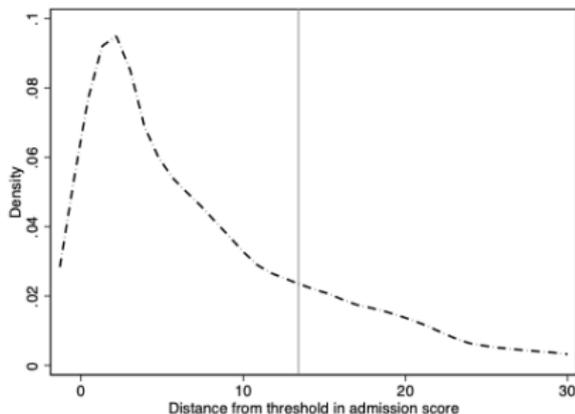
(b) % from Schools in Bottom 20% of Achievement

- Share of low-SES students in elite schools doubles with no demand effect
 ⇒ High-SES students decreasing demand for elite

Causal Effect of Elite Admission for Low-SES Students



(a) Graduation on Time in Previous Cohorts



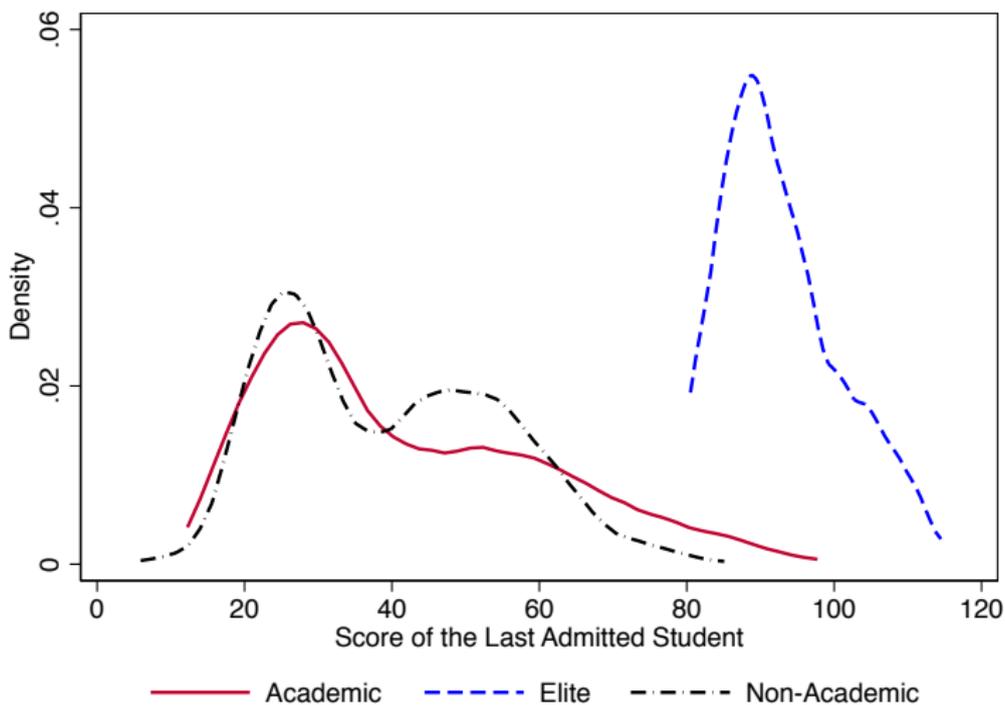
(b) Applicants Admitted under Policy Counterfactual

- (Marginal) elite admission \downarrow by 11-12 p.p. the graduation rate for low-SES
- RD sample is representative for policy counterfactual

Wrapping Up

- Biased perceptions about ability may distort schooling decisions
- Providing personalized information via test scores
 - Changes high-school track choices and placement
 - Increases on-time graduation by better aligning skills with educational choices
- Equilibrium effects may temper positive impact for low-SES applicants
 - Positive average impact on student ex-ante welfare
 - Information may not be enough to improve outcomes for targeted students

Distribution of Cutoff Scores

[← Back](#)

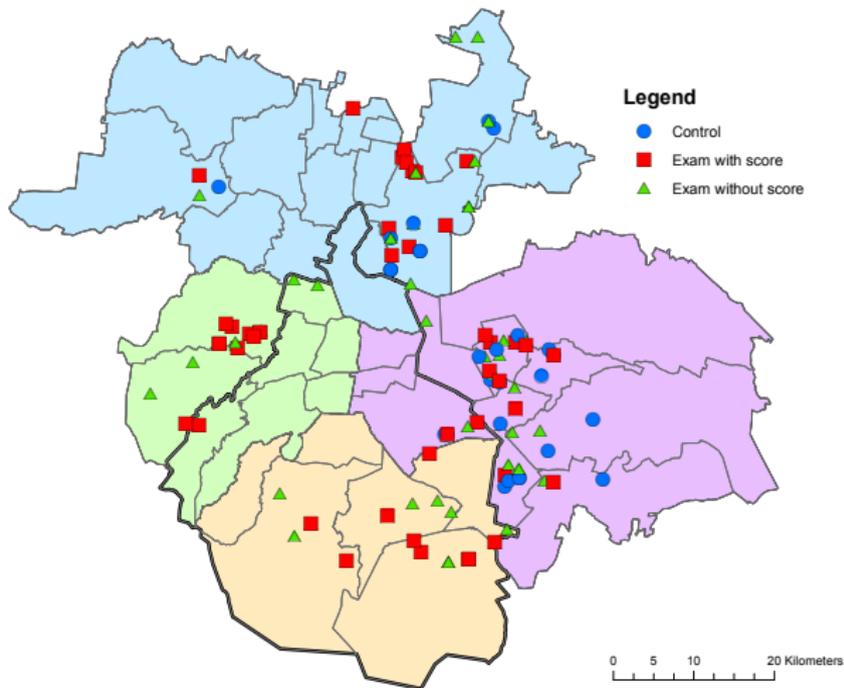
Applicants' Characteristics in the Population and in the RCT Sample

[◀ Back](#)

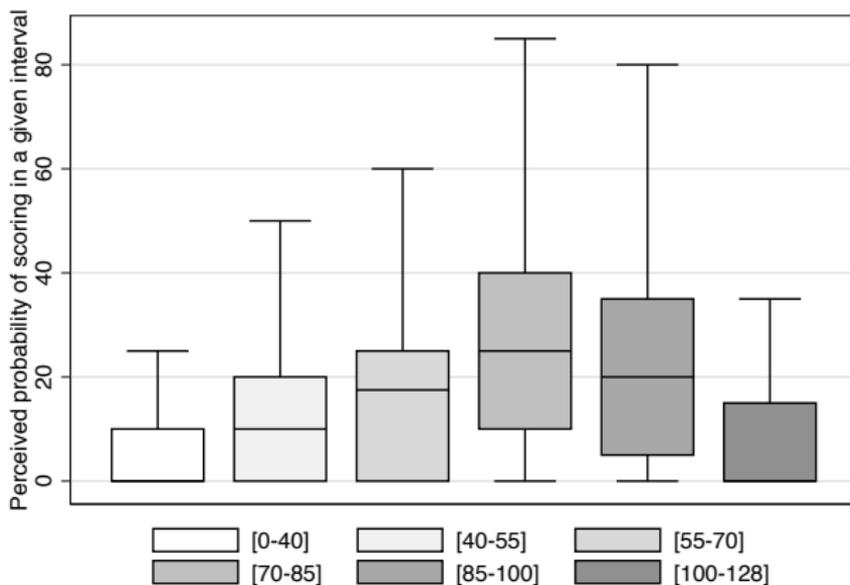
	All COMIPEMS	Experiment	Difference [<i>p</i> -values]
Admission exam score	69.506 (20.705)	65.400 (19.401)	4.107 [0.000]
Grade Point Average in middle school (GPA)	8.058 (0.871)	8.119 (0.846)	-0.061 [0.001]
Has some disabilities (1=yes)	0.118 (0.323)	0.145 (0.352)	-0.027 [0.000]
Scholarship in middle school (1=yes)	0.116 (0.320)	0.110 (0.313)	0.006 [0.401]
Indigenous	0.041 (0.198)	0.093 (0.290)	-0.052 [0.000]
Plans to go to college (1=yes)	0.662 (0.473)	0.670 (0.470)	-0.008 [0.378]
One parent with at least tertiary education (1=yes)	0.236 (0.425)	0.147 (0.354)	0.089 [0.000]
Average math score in middle school (z-score)	0.000 (1.000)	-0.208 (0.712)	0.208 [0.000]
Neighborhood poverty index (z-score)	0.000 (1.000)	1.504 (0.494)	-1.504 [0.000]
Observations	284,412	2,493	

NOTE: The first two columns report means and standard deviation (in parentheses) of individual characteristics between the overall population of applicants and the experimental sample. The third column displays mean differences and the associated *p*-values (in brackets) for the null hypothesis of equal means. The observations in the 'All COMIPEMS' column comprise all the applicants in the year 2014 who were eligible to be assigned through the matching algorithm. The observations in the 'Experiment' column comprise the evaluation sample of the randomized information intervention.

Schools in the RCT Sample

[← Back](#)

Distribution of beliefs over discretized support [◀ Back](#)



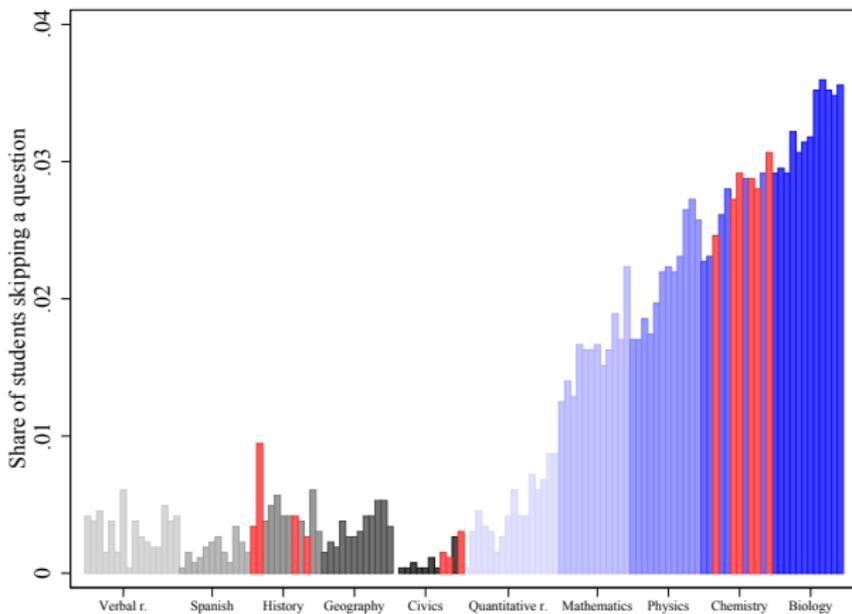
excludes outside values

- Assuming a uniform distribution within each interval of the score, the expected scores are constructed as the summation over intervals of the product of the mid-point of the bin and the probability assigned by the student to that bin

Descriptive Statistics on the Score in Mock Test [← Back](#)

- The average number of skipped questions in our mock exam is only 1.4 out of 128, and more than 80% did not leave any question unanswered
 - Skipping patterns seem more consistent with binding time constraints rather than lack of seriousness [Skipping](#)
- Correlation between performance in the mock exam and the admission exam is 0.82
- Correlation between middle-school GPA and the admission exam score is 0.48
- The score in the mock exam predicts high-school outcomes, even after controlling for middle school GPA
 - One SD increase is associated with a 2.6 pp. increase in the probability of graduating from high school on time.

Figure: Average Skipping Patterns in the Mock Exam



Descriptive Statistics on High-School Trajectories

[← Back](#)

- About 80 percent of the students in the control group enroll in the school they were assigned in the first placement round.
- Conditional on enrollment, only 56 percent graduate on time from high school
 - 66% in academic and 45% in non academic (vocational/technical)
 - These figures are not peculiar to the experimental sample and they are strong indicators of mismatch between schools and students

Gap Between Expected and Realized Score ◀ Back

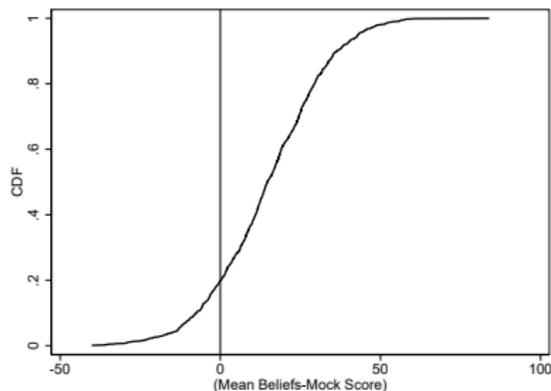
	Abs(Gap)			
	All Sample	By Performance Index		
		Tercile I	Tercile II	Tercile III
Treatment	-6.809 [0.000] {0.001}	-8.341 [0.000] {0.001}	-6.827 [0.000] {0.001}	-5.170 [0.000] {0.001}
Mean Control	18.8	23.8	18.0	14.6
Number of Observations	2178	683	740	755
Number of Clusters	90	90	90	87
R-squared	0.100	0.131	0.132	0.114

NOTE: All specifications include a set of dummy variables which correspond to the randomization strata, pre-determined characteristics (sex, characteristics of the school of origin, previous experience with practice exams providing feedback, aspirations to attend college, an index of personality traits, an index of parental characteristics, and a household asset index), and indicator variables for whether each of the covariates has missing data. The dependent variable “Abs(Gap)” is the absolute value of the difference between the expected score (as captured in the survey) and the score in the mock exam. p -values reported in brackets refer to the conventional asymptotic standard errors while those reported in curly brackets are adjusted for testing each null hypothesis across multiple outcomes through the step-wise procedure described in [Romano & Wolf \(2005a,b, 2016\)](#). All inference procedures take into account the clustering of error terms at the middle school level and the block randomization design.

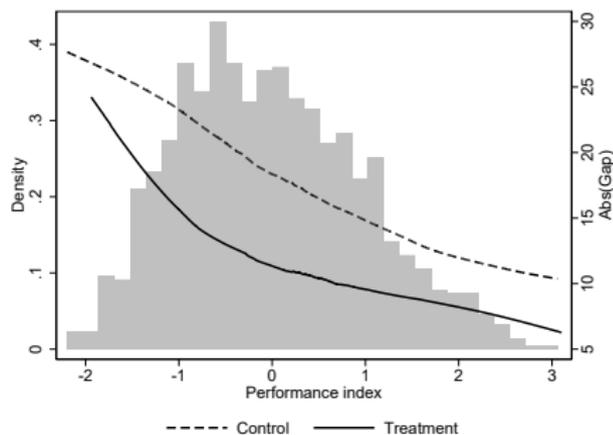
Subjective Expectations about Test Performance

[← Back](#)

(a) CDF of the Gap (Control Group)



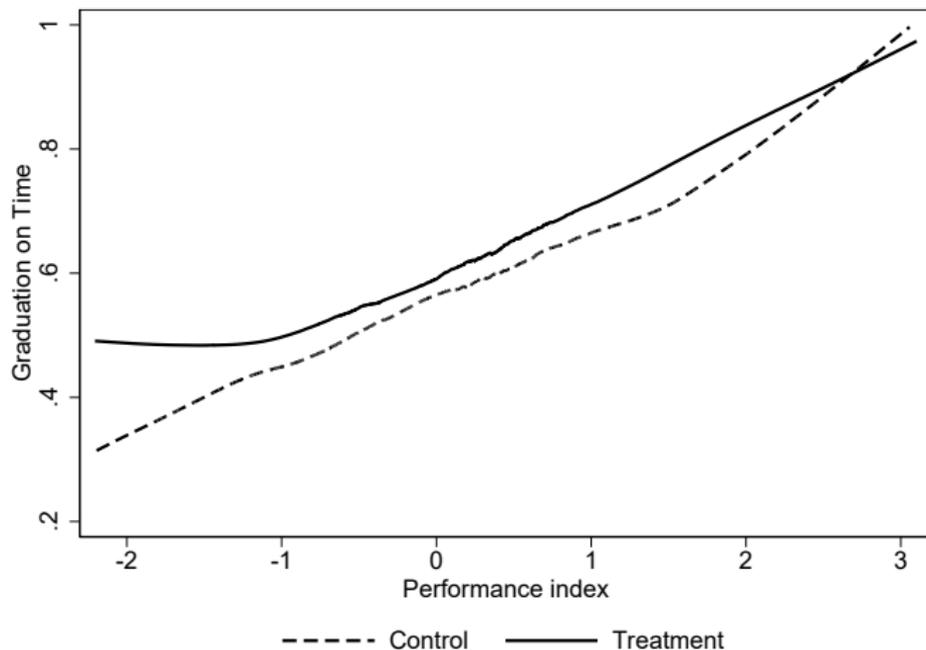
(b) Absolute Gap by Ability



School Choices and Placement Outcomes [◀ Back](#)

	Non-academic Schools		Academic Schools		Elite Schools	
	Share in ROL	Placement	Share in ROL	Placement	Share in ROL	Placement
Treatment	0.002 [0.918] {0.999}	0.047 [0.074] {0.195}	-0.001 [0.923] {0.999}	-0.045 [0.076] {0.195}	-0.000 [0.986] {0.999}	-0.002 [0.845] {0.998}
Performance Index	-0.031 [0.002] {0.007}	-0.079 [0.000] {0.001}	-0.054 [0.000] {0.001}	-0.086 [0.000] {0.001}	0.084 [0.000] {0.001}	0.165 [0.000] {0.001}
Treat × Performance Index	-0.032 [0.010] {0.030}	-0.065 [0.015] {0.041}	0.030 [0.008] {0.023}	0.041 [0.046] {0.107}	0.002 [0.865] {0.999}	0.024 [0.243] {0.549}
Mean Control	0.363	0.401	0.328	0.369	0.309	0.114
Number of Observations	2493	2493	2493	2493	2493	2493
Number of Clusters	90	90	90	90	90	90
R-squared	0.154	0.101	0.129	0.061	0.264	0.335

High-School Graduation on Time and Academic Achievement

[◀ Back](#)

Education Outcomes in Upper Secondary ◀ Back

	Enrollment	Dropout in 1st year	Graduation on Time
Treatment	-0.005 [0.751] {0.697}	0.015 [0.500] {0.614}	0.040 [0.075] {0.055}
Mean Control	0.8	0.2	0.6
Number of Observations	2492	2023	1888
Share of Missing Data	0.000	0.189	0.243
Lee lower bound		0.008	0.035
Lee upper bound		0.011	0.060
R-squared	0.064	0.167	0.127

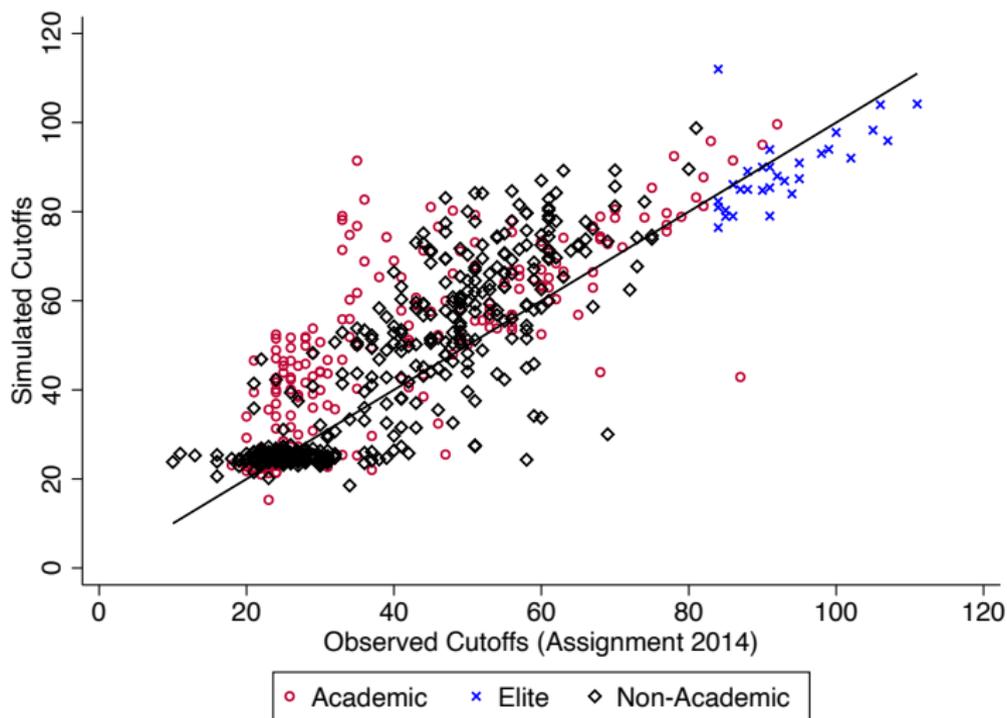
NOTE: All specifications include a set of dummy variables that correspond to the institution in which the student was placed, pre-determined characteristics (sex, characteristics of the school of origin, previous experience with practice exams providing feedback, aspirations to attend college, an index of personality traits, an index of parental characteristics, and a household asset index), and indicator variables for whether each of the covariates has missing data. The dependent variable “Enrollment” denotes an indicator variable that is equal to one if students enroll in the high school programs they were assigned to, and zero otherwise. The dependent variables “Dropout, 1st year” captures whether the student stopped attending classes or actively dropped out of school, conditional on enrollment. The dependent variable “Graduation on Time” denotes an indicator variable that is equal to one if the student successfully completes the high school programs three years after enrolling in tenth grade, and zero otherwise. Lee bounds (Lee 2009) are reported at the bottom of the table in order to account for potentially non-random sample attrition. p -values reported in brackets refer to the conventional asymptotic standard errors, while those reported in curly brackets are adjusted for by testing each null hypothesis across multiple outcomes through the step-wise procedure, as described in Romano & Wolf (2005a,b, 2016). All inference procedures take into account clustering of the error terms at the middle school level and the block randomization design.

Model Fit on Assignment Outcomes ◀ Back

	Data	Model	Difference
Applied in the system (1=yes)	1.00	0.99	-0.01
Assigned in the system (1=yes)	0.87	0.89	0.02
<u>Assigned in:</u>			
Vocational schools	0.14	0.11	-0.03
Technical schools	0.26	0.26	-0.00
Academic schools	0.38	0.41	0.03
Elite schools	0.23	0.22	-0.00
Selectivity (z-cutoff score)	0.75	0.74	-0.01
Academic (above-median selectivity)	0.50	0.50	-0.01
Academic (below-median selectivity)	0.10	0.13	0.03
Non-academic (above-median selectivity)	0.25	0.24	-0.01
Non-academic (below-median selectivity)	0.15	0.13	-0.02

NOTE: The moments displayed in the first column are computed from the data of the assignment mechanism in the year 2014 (see Section 2). The moments displayed in the second column are computed by running the Serial Dictatorship algorithm that is in place for the COMIPEMS system, using the simulated rank-ordered lists from the estimates reported in the fourth column in Appendix Table B.5, the individual scores in the admission exam, and the school capacities as inputs.

Model Fit on Schools' Cutoff Scores

[← Back](#)

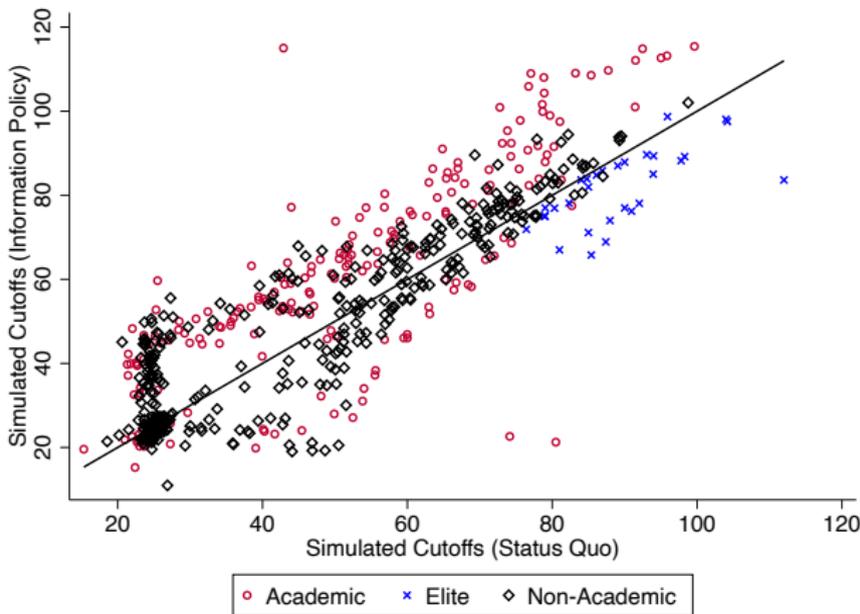
Feedback Provision at Scale: Assignment Outcomes

[← Back](#)

	Status Quo	Information Intervention	Difference
Applied in the system (1=yes)	0.99	0.99	0.00
Assigned in the system (1=yes)	0.89	0.91	0.02
Rank of assigned school	6.41	5.43	-0.98
Assigned in top choice	0.16	0.25	0.09
Assigned in elite schools	0.22	0.22	0.00
Assigned in academic schools	0.41	0.40	-0.01
Assigned in non-academic schools	0.37	0.38	0.01

- 1 Small increase in the share of assigned students
- 2 Average applicant is placed in more desirable school (Rank=5.4 Vs. 6.4)
- 3 ↑ Share of students assigned to their most preferred option from 16 to 25%
- 4 Aggregate sorting patterns across tracks unchanged

Feedback Provision at Scale: Cutoff Scores

[← Back](#)

- Movement in cutoff scores explains lack of aggregate effect on sorting

Average Marginal Effects for Selected School-Student Match Parameters

[◀ Back Demand Effect](#)

	Control Sample		Treatment Sample	
	Coefficient	Std. Error	Coefficient	Std. Error
Academic \times GPA	0.0005	0.0016	0.0036	0.0020
Academic \times Poverty	-0.0061	0.0024	-0.0074	0.0036
Elite \times GPA	0.0053	0.0046	0.0046	0.0057
Elite \times Poverty	0.0064	0.0081	0.0348	0.0099
Selectivity \times GPA	0.0027	0.0012	0.0039	0.0014
Selectivity \times Poverty	-0.0042	0.0019	-0.0058	0.0026

NOTE: This table depicts the estimated average marginal effects for selected school-student match parameters, which are computed using the model estimates shown in Table B.5. These coefficients show the change in the conditional probability that student i chooses a school j with a given characteristic (academic, elite, and the degree of selectivity), resulting from a one-unit increase in the individual covariates (GPA and Poverty). Standard errors are computed using the delta method.