Randomized Control Trials and Policy Evaluation

Matteo Bobba
matteo.bobba@tse-fr.eu
Office: T.353

Toulouse School of Economics (TSE)

M2 PPD-EEP-EEE

Spring 2026

Part 4: Scaling-up RCTs

1. Enhancing Human Capital at Scale: A Case Study on Scaling (JPE, 2025)

   $\Rightarrow$ Francesco Agostinelli (Penn), Ciro Avitabile (WB), Matteo Bobba (TSE)

2. Perceived Ability and School Choices: Experimental Evidence and Scale-up Effects

   $\Rightarrow$ M. Bobba (TSE), Veronica Frisancho (CAF), Marco Pariguana (Edinburgh)

Enhancing Human Capital at Scale: A Case Study on Scaling

Motivation

- Interest in the ability of RCTs to inform policy decisions

  $\Rightarrow$ Large effect in RCT $\rightarrow$ voltage drop at scale (List, 2022)

- The devil is in the details (of the implementation)

  $\Rightarrow$ Small changes in intervention often translate into large differences in ATE

- We provide a case study on scaling educational interventions

  $\Rightarrow$ Two RCTs + government implementation of a mentoring program

Context: CONAFE Schools in Rural Mexico

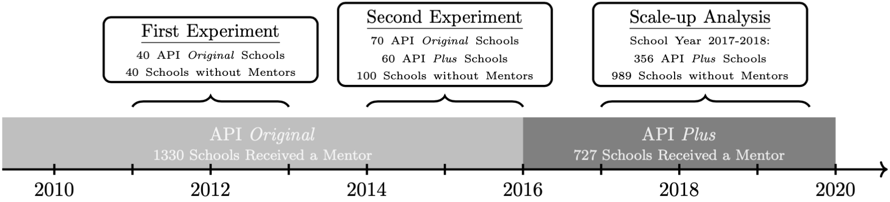
- Government agency provides schooling services to poor/remote villages
  - ⇒ Villages with $<$ 2,500 residents
  - ⇒ Small multi-grade schools (10-15 students)
  - ⇒ Average yearly rate of school closures is 11%
- Community-based schooling model
  - ⇒ Community instructors with little training and high turnover
  - ⇒ Parents organize local associations aimed at promoting education

The Mentoring Program: API

- Meetings with parents (home and school)
- One-on-one tutoring for the six weakest students
- Pedagogical support to instructors
- API Original
  - $\Rightarrow$ Mentors are selected among college students/graduates
  - $\Rightarrow$ Initial training + bi-monthly workshops on curricular knowledge
- API Plus
  - $\Rightarrow$ Extra-training + bi-monthly workshops to improve parenting skills
  - $\Rightarrow$ Peer-to-peer sessions (18 hours): sharing effective practices

Research Design: Timeline and Sampling

Figure 1: The Mentoring Program in Chiapas and the Different Study Samples



- Programâs 2-year rotation cycle $\rightarrow \approx$360 mentors for each program cycle

  $\Rightarrow$ Among the API Plus schools in 2017â2018, 86 were in second experiment

Sample Representativeness

Table 1: Differences Across Populations

	Panel A: School Characteristics				
	All Chiapas Mean (SD)	First Experiment Mean (SD)	Second Experiment Mean (SD)	Chiapas vs. Experiment 1 Mean Difference [p-value]	Chiapas vs. Experiment 2 Mean Difference [p-value]
Average Test Score (Spanish)	424.503	399.116	431.340	-25.387	6.837
	(56.466)	(32.631)	(60.810)	[0.000]	[0.139]
Average Test Score (Math)	414.921	379.165	421.333	-35.756	6.412
	(75.300)	(45.339)	(80.895)	[0.000]	[0.297]
Number of Students	14.049	15.507	15.009	1.458	0.960
	(8.468)	(8.781)	(6.053)	[0.175]	[0.037]
Number of Teachers	1.231	1.333	1.217	0.102	-0.014
	(0.467)	(0.505)	(0.413)	[0.099]	[0.638]
Share Over-aged Students	0.349	0.230	0.324	-0.119	-0.025
	(0.797)	(0.552)	(0.659)	[0.088]	[0.610]
Number of Schools	1,523	80	230	1,603	1,753

- First experiment is not fully representative
- Sample of the second experiment resembles target population in Chiapas

API Original vs. Plus: Experimental Evidence

	Survey-Based Test Scores				Administrative Records	
	Reading (1)	Math (2)	Socio-Emotional (3)	Overall Index (4)	Enroll (5)	Secondary (6)
API Original	.126	.056	.071	.126	.073	.081
	[.104]	[.455]	[.418]	[.182]	[.255]	[.519]
	{.134}	{.486}	{.446}	{.222}	{.281}	{.567}
	(.150)	(.558)	(.558)	(.240)	(.311)	(.478)
API Plus	.315	.237	.199	.368	.124	.298
	[.001]	[.008]	[.022]	[.001]	[.074]	[.030]
	{.001}	{.014}	{.032}	{.001}	{.089}	{.052}
	(.001)	(.005)	(.011)	(.001)	(.030)	(.030)
API Original = API Plus	[.043]	[.043]	[.178]	[.021]	[.469]	[.134]
	{.077}	{.112}	{.221}	{.024}	{.568}	{.230}
	(.045)	(.045)	(.100)	(.024)	(.372)	(.156)
Schools (no.)	224	224	224	224	182	76
Observations	1,044	1,044	1,045	1,045	468	106

API Plus Scale-up: Empirical Strategy

- Program assignment follows a priority-based mechanism across communities
- We leverage this variability to assess the impact of API Plus at scale

$$Y_j = \alpha_0 + \alpha_1 \text{Plus}_j + \boldsymbol{\delta' X_j} + \epsilon_j$$

$\Rightarrow$ $\boldsymbol{X_j}$: indicators for the four priority criteria + hostile events + prior API original

- Standard CIA: $E(\epsilon_j \mid \text{Plus}_j, \boldsymbol{X_j}) = E(\epsilon_j \mid \boldsymbol{X_j})$

  $\Rightarrow$ Observables are balanced for Plus $\in \{0, 1\}$ after conditioning on $\boldsymbol{X_j}$

  $\Rightarrow$ Placebos using achievement scores prior to API Plus at scale

API Plus Scale-up: Results

	Nonexperimental Schools		Experimental Schools	
	Enroll Secondary (1)	Child Literacy (2)	Enroll Secondary (3)	Child Literacy (4)
API Plus	.056	.028	.091	.035
	[.010]	[.012]	[.022]	[.078]
	{.011}	{.014}	{.021}	{.075}
	(.020)	(.020)	(.041)	(.075)
Schools (no.)	1,161	1,161	184	184

- Treatment effect on enroll is in line with RCT evidence (e.g., $\uparrow$ 9.1pp vs 12pp)
- Reduction of illiteracy rates by 20% with respect to the sample average

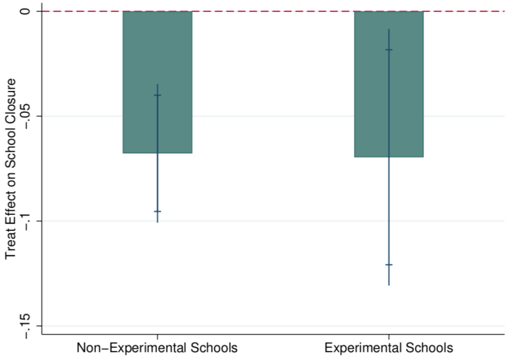
Takeaways

- API Original is ineffective
  - ⇒ Joint null across both experiments has a p-value $=$ 0.460
- API Plus modality has large effects on children outcomes
  - ⇒ Across different samples of schools under the same implementation regime
  - ⇒ Across different regimes (RCT and gov) for the same sample
  - ⇒ Joint null is highly significant (p-values < 0.001)

The Threat of Voltage Drop in the Scale-up

- Robust elements of the design of the evaluation
  - ⇒ Program impacts (Original & Plus) across multiple samples and situations
  - ⇒ Representative sample of communities in Chiapas
  - ⇒ Experiments designed within existing infrastructure of the program at scale
  - ⇒ Randomization at large unit level, encompassing community-level spillovers
- Remaining Challenges
  - ⇒ Quantity/Quality of Mentoring: No evidence of major changes
  - ⇒ School closures: Only two schools closed in the experimental sample

The Impact of the API Plus Program on School Closures



- The scale-up implementation drastically reduces incidence of school closures

Pathways to Scale

- Parents play an important role in the school community
- Mentors with enhanced training trigger parental engagement
  - $\Rightarrow$ Parent/mentor and parent/child interactions
  - $\Rightarrow$ No differential effect of remedial education sessions or pedagogical support
- Community-level parental engagement may preserve “voltageââ when scaling

Original vs Plus: Parental Investment

	Engage at School (1)	Manage School Resources (2)	Engage With Child (3)	Overall Index (4)
	A. First Experiment			
API Original	.198	−.135	.149	.101
	[.259]	[.415]	[.399]	[.580]
	{.261}	{.422}	{.399}	{.578}
	(.338)	(.511)	(.511)	(.511)
Schools (no.)	73	73	73	73
Observations	208	208	208	208
	B. Second Experiment			
API Original	−.188	−.124	.167	−.034
	[.049]	[.176]	[.015]	[.684]
	{.070}	{.216}	{.030}	{.709}
	(.058)	(.197)	(.015)	(.630)
API Plus	.217	.087	.353	.359
	[.034]	[.344]	[.001]	[.001]
	{.047}	{.393}	{.001}	{.001}
	(.037)	(.247)	(.000)	(.001)
API Original = API Plus	[.001]	[.056]	[.029]	[.001]
	{.000}	{.058}	{.171}	{.001}
	(.002)	(.036)	(.036)	(.001)
Schools (no.)	224	224	224	224
Observations	1,045	1,045	1,045	1,045

Original vs Plus: Channels

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	A. Parents and Mentors Interactions (as Reported by Parents)						
	Quantity (Last 60 Days)			Quality			
	Meetings	Visits	Index	Inform About Child	Advise About Child	Index	
API Plus	1.039	.726	.362	.102	.100	.251	
	[.147]	[.125]	[.062]	[.057]	[.034]	[.040]	
	{.194}	{.171}	{.094}	{.097}	{.056}	{.070}	
	(.194)	(.194)	(.100)	(.078)	(.078)	(.078)	
Observations	482	491	504	354	353	357	
	B. Parenting Styles Promoted by Mentors (as Reported by Mentors)						
	Educative Style			Emotional Style			
	Communication	Learning	Index	Share Feelings	Self-Knowledge	Manage Transitions	Index
API Plus	.178	.168	.494	.049	.030	.142	.194
	[.038]	[.077]	[.018]	[.627]	[.756]	[.123]	[.312]
	{.049}	{.092}	{.024}	{.637}	{.749}	{.118}	{.315}
	(.070)	(.070)	(.040)	(.846)	(.846)	(.295)	(.542)
Observations	107	107	107	107	107	107	107

Parents as Means of Scalability

	Outcome: School Closures		
	First Experiment (1)	Second Experiment (2)	Second Experiment, IV (3)
API Original	.063	−.031	−.031
	[.225]	[.396]	[.410]
API Plus		−.083	
		[.030]	
Overall parental engagement			−.217
			[.021]
Observations	73	224	1,045
Clusters			224
F-statistic (excl. instrument)			13.833

- $+0.1$sd of parental engagement $\uparrow$ 2.2pp probability that school remains open

What Drives the Success of the Program at Scale?

- Active parental involvement prevented voltage drop
  - ⇒ Qualitative evidence suggests parents ensure continuity in schooling activities
  - ⇒ Parental responses are shown to prevent schools from closing
- Implementation matters for scaling-up RCTs
  - ⇒ Promoting parental engagement in education can go a long way
  - ⇒ This may hamper or sustain the scalability of schooling interventions

Perceived Ability and School Choices: Experimental Evidence and Scale-up

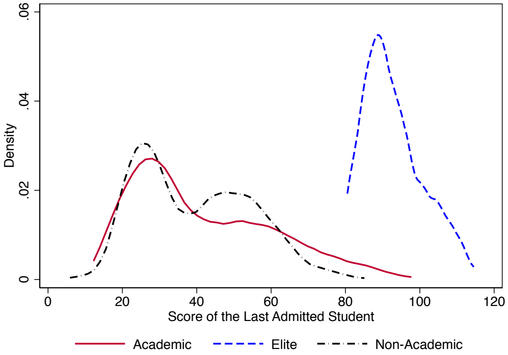
Motivation

- Human capital investments rely on expected individual-specific net returns
- How does information about own academic skills shape choices/outcomes?
  - $\Rightarrow$ RCT provides sub-set of students with feedback on their academic skills
  - $\Rightarrow$ Quantify impacts on beliefs, school choice/placement and educ. outcomes
- What are the effects of such information intervention at scale?
  - $\Rightarrow$ Cheap $\neq$ scalable: at scale you may have equilibrium effects
  - $\Rightarrow$ A model of school choice + outcomes to simulate large-scale implementation

Context: Centralized School Assignment in Mexico City

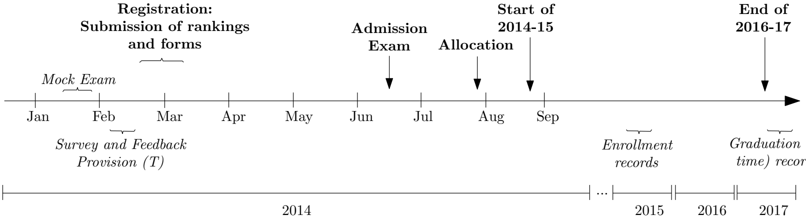
- Centralized admission system for public high schools
  - $\Rightarrow$ $\approx$ 300k applicants and more than 600 schools
  - $\Rightarrow$ Rank-ordered lists (ROLs) submitted before students know admission score
- High-school tracks: General and Technical/Vocational (non-acad.)
  - $\Rightarrow$ General track is oriented toward tertiary education (acad.)
  - $\Rightarrow$ 32 (elite) schools are affiliated with two prestigious higher educ. institutions

Distribution of Cut-off Scores



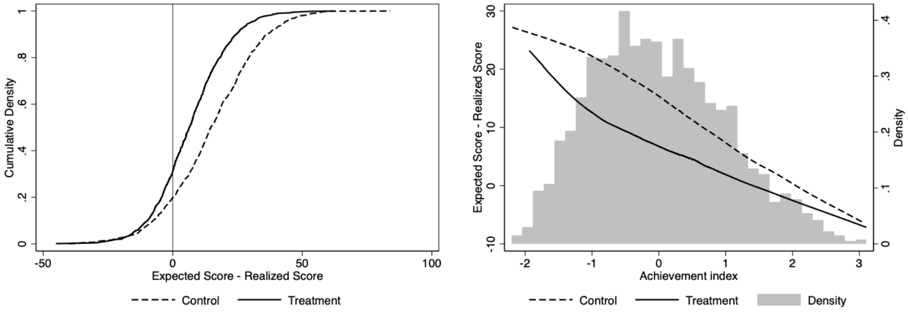
- Large overlap in cut-off scores across acad. and non-acad tracks
- Elite schools clearly stand out in terms of selectivity/quality of peers

Experimental Design



- RCT sample consists of 9th graders in relatively disadvantaged city blocks
  - $\Rightarrow$ $\approx$ 2,500 applicants in 90 schools ($< 1\%$ of total applicants)
  - $\Rightarrow$ Corr(mock score, actual test score) $=$ 0.82
  - $\Rightarrow$ Both scores are strong predictors of later education outcomes

Performance Feedback and Belief Updating



(a) Cumulative Density of the Gap

(b) Gap and Academic Achievement

- Performance feedback more than halves the average perception gap

⇒ More results on belief updating in Bobba and Frisancho (2022)

Performance Feedback and Assignment Outcomes

	Non-Academic	Academic	Elite
Treatment	0.046	-0.044	-0.002
	[0.077]	[0.078]	[0.861]
	{0.110}	{0.110}	{0.842}
Achievement index	-0.079	-0.086	0.165
	[0.000]	[0.000]	[0.000]
	{0.002}	{0.001}	{0.001}
Treatment × Achievement index	-0.065	0.041	0.024
	[0.015]	[0.045]	[0.247]
	{0.020}	{0.065}	{0.253}
Mean Control	0.453	0.418	0.129
Number of Observations	2493	2493	2493
Number of Clusters	90	90	90
R-squared	0.100	0.061	0.336

- Placement in non-academic ↑ by 5 pp and ↓ by 4 pp in academic (noisy)
  - ⇒ Feedback alters the skill composition across high-school tracks
  - ⇒ No effect on admission exam or other application outcomes

Performance Feedback and High School Outcomes

	Enrollment	Dropout 1st year	Graduation on Time
Treatment	-0.003	0.012	0.022
	[0.789]	[0.668]	[0.252]
	{0.936}	{0.920}	{0.497}
Achievement index	0.068	-0.095	0.138
	[0.000]	[0.001]	[0.000]
	{0.001}	{0.003}	{0.001}
Treatment × Achievement index	-0.021	-0.006	-0.032
	[0.352]	[0.807]	[0.088]
	{0.590}	{0.936}	{0.198}
Mean Control	0.813	0.248	0.447
Number of Observations	2493	2024	2358
R-squared	0.045	0.076	0.090

- One-std dev $\downarrow$ in achievement $\uparrow$ the on-time graduation by 5.4 pp (noisy)

  $\Rightarrow$ Effect not driven by choosing/being assigned to easier-to-graduate schools

Experimental Evidence on Performance Feedback

- Feedback shifts individual belief distributions about student skills

  $\Rightarrow$ Both first and second moments

- Feedback is consequential for sorting patterns across high-school tracks

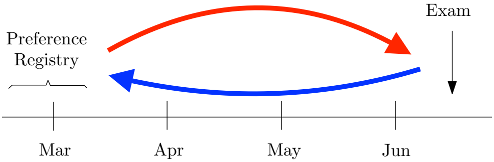
  $\Rightarrow$ Average and composition effects on the acad/non-acad. margin

- Feedback boosts high-school graduation for lower performing students

  $\Rightarrow$ Information intervention contributes to “leveling the playing field”

Scaling-up the Information Experiment

- Would this evidence replicate in a large-scale implementation? E.g.:



⇒ Or universal mock exam with score delivered through application platform

- Key threat to scale-up with fixed school capacities: congestion externalities

⇒ Aggregate shift in the demand-side may induce displacement effect

Ex Ante Policy Evaluation: A Model-based Approach

- We estimate a model of school choice to predict aggregate sorting patterns

  $\Rightarrow$ Match$(T) = \mu(u(\widehat{\theta^T}),$ Priorities, School Capacities), where $T \in \{0, 1\}$

- We estimate a value-added model to map sorting into educ outcomes

  $\Rightarrow$ School VA heterogeneous across students + school-level peer effects

School Choice Model

- Indirect utility that student i gets from attending school j/college s:

$$u_{ij} = \underbrace{\alpha_{s(j)} + \beta'_{s(j)}\hat{a}x_i}_{\text{Net Returns}} + \underbrace{\gamma' x_i d_{ij}}_{\text{Commuting Costs}} + \underbrace{\rho x_i \hat{a} c_j}_{\text{Select/Peers}} + \epsilon_{ij}$$

- $\epsilon_{ij} \sim$ type-I extreme value
- Outside option: $u_{i0} = \epsilon_{i0}$
- Tuition fees are captured by the $\alpha_{s(j)}$
- Conditional on $\boldsymbol{x_i}$, d_{ij} is assumed orthogonal to preference shock ϵ_{ij}

Estimating Preferences

- School rankings may deviate from true preference orderings
  - $\Rightarrow$ ROLs are truncated at 20 schools
  - $\Rightarrow$ Possible (strategic) mistakes in applications
- Realized match is likely ex-post stable (large market)

$$\Pr(D_i = j) = \frac{\exp\left(\alpha_{s(j)} + \boldsymbol{\beta}'_{\boldsymbol{s(j)}}\boldsymbol{x_i} + \boldsymbol{\gamma}'\boldsymbol{x_i}d_{ij} + \boldsymbol{\rho}'\boldsymbol{x_i}c_j\right)}{\sum_{l\in\Omega_i}\exp\left(\alpha_{s(l)} + \boldsymbol{\beta}'_{\boldsymbol{s(l)}}\boldsymbol{x_i} + \boldsymbol{\gamma}'\boldsymbol{x_i}d_{il} + \boldsymbol{\rho}'\boldsymbol{x_i}c_l\right)}, \Omega_i = \{j : s_i \geq \kappa_j\}.$$

- ML estimates show stark differences across treated and control applicants
  - $\Rightarrow$ Higher-SES with feedback attach a more negative value to an elite school
  - $\Rightarrow$ Consistent with belief updating in Bobba and Frisancho (2022)

Out-of-sample Sorting Predictions

- Matching eq. using estimated preferences for the control group replicate data

	Very Low SES		Low SES		Middle SES		High SES	
	Data	Model	Data	Model	Data	Model	Data	Model
Applied in the system (1=yes)	1.00	0.97	1.00	0.99	1.00	0.99	1.00	1.00
Assigned in the system (1=yes)	0.91	0.91	0.88	0.92	0.86	0.93	0.84	0.94
Non-Academic schools, vocational track	0.16	0.18	0.14	0.13	0.13	0.10	0.10	0.08
Non-Academic schools, technical track	0.30	0.27	0.27	0.27	0.25	0.25	0.22	0.23
Academic, above-median selectivity	0.23	0.20	0.30	0.28	0.30	0.31	0.32	0.33
Academic, below-median selectivity	0.17	0.21	0.09	0.13	0.04	0.08	0.03	0.04
Elite schools	0.13	0.14	0.20	0.18	0.28	0.26	0.34	0.32
Selectivity (z-cutoff score)	0.32	0.24	0.65	0.56	0.96	0.90	1.21	1.15

$\Rightarrow$ Correlation between observed and simulated cut-offs is 0.88

Linking Sorting to Education Outcomes

- Potential outcome of student i if she is matched to school j:

$$Y_{ij} = \underbrace{\delta_{s(j)}}_{\text{Avg Effect of } s} + \underbrace{\boldsymbol{\gamma'_{s(j)} x_i}}_{\text{Match Effect of } s \text{ and } x_i} + \underbrace{\boldsymbol{\lambda' \bar{x}_j}}_{\text{Peers in } j} + \nu_{ij}$$

- School placement depends on school rankings and exam scores
  - $\Rightarrow$ We add ROL fixed effects: $r_k = \mathbb{I}[R_i = k]$
  - $\Rightarrow$ Variation in ν_{ij} due to idiosyncratic differences in admission score

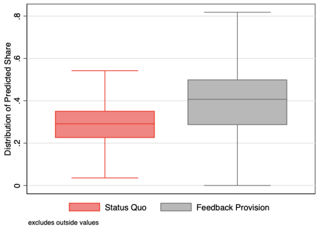
- Attending an elite school decreases on-time graduation rates by 18 pp.
  - $\Rightarrow$ Match effects are much less important than average effect
  - $\Rightarrow$ School-level peer effects also play a role

Performance Feedback and Matching Outcomes

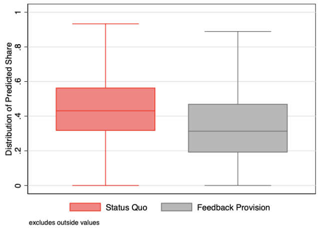
	Status Quo	Information Intervention	Difference
Applied in the system (1=yes)	0.99	0.99	0.00
Assigned in the system (1=yes)	0.89	0.91	0.02
Rank of assigned school	6.41	5.43	-0.98
Assigned in top choice	0.16	0.25	0.09
Assigned in elite schools	0.22	0.22	0.00
Assigned in academic schools	0.41	0.40	-0.01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

- Info provision enhances the ex-ante efficiency of the matching equilibrium
  - $\Rightarrow$ Share of students assigned to most preferred option $\uparrow$ by 9 pp.
  - $\Rightarrow$ Share of students assigned through the algorithm $\uparrow$ by 2 pp.
  - $\Rightarrow$ No changes at the extensive margin or in average sorting across tracks

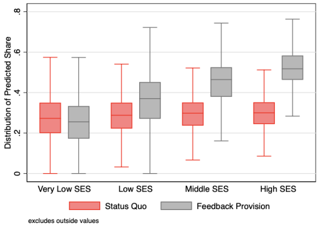
Performance Feedback and School Choices



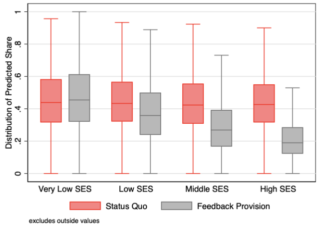
(a) Aggregate Shares of Academic Schools



(b) Aggregate Shares of Elite Schools



(c) Shares of Academic Schools by SES

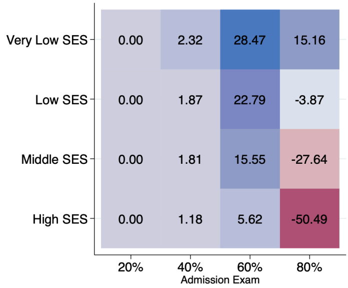


(d) Shares of Elite Schools by SES

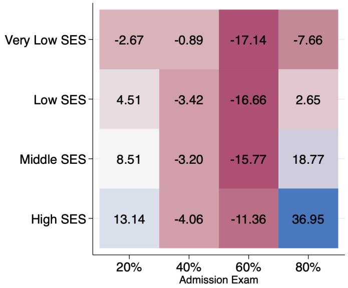
Performance Feedback and School Choices

- Overall $\uparrow$ in demand for acad. schools and a symmetric $\downarrow$ for elite schools
  - $\Rightarrow$ Changes in equilibrium cutoff scores reflect this shift in aggregate demand
  - $\Rightarrow$ And can explain the muted effect on average sorting across tracks
- The bulk of the changes in school preferences accrue to high-SES applicants
  - $\Rightarrow$ Negative gradient wrt SES for treated applicants in value of elite school
  - $\Rightarrow$ Low-SES are, on average, unresponsive to the intervention
- Scaling up feedback would alter the SES composition of elite schools
  - $\Rightarrow$ $\downarrow$ demand for elite among high-SES leaves open seats for low-SES

Performance Feedback and High-School Assignment



(a) Elite Schools

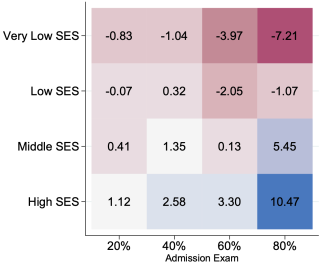


(b) Academic Schools

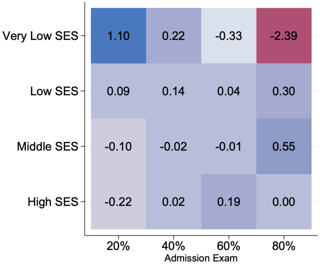
- Information intervention ↑ low-SES in elite schools by more than 20 pp.

  ⇒ RCT documents no effect of feedback provision on the assignment in elite

Performance Feedback and High-School Graduation



(a) Full Scale Implementation



(b) Feedback Targeted to Very Low SES

- Crowd-in effect at elite ↓ high-school graduation for low-SES by 4-7 pp.
  - ⇒ Positive gradient wrt achievement for high-SES mirrors assignment in acad
  - ⇒ Targeted intervention to low-SES is consistent with RCT evidence

Conclusion

- RCT shows that performance feedback improves student-school matches
  - $\Rightarrow$ Higher graduation rates three years post-assignment
- Scale-up simulations reveal that congestion externalities overturn that effect
  - $\Rightarrow$ Scaling may require additional signals targeted at low-SES students
- Info provision in centralized education markets may enhance efficiency
  - $\Rightarrow$ Distributional consequences are far more nuanced