Randomized Control Trials and Policy Evaluation

Matteo Bobba
matteo.bobba@tse-fr.eu
Office: T.363

Toulouse School of Economics (TSE)

M2 PPD/ERNA/EEE, Winter 2024

Enhancing Human Capital in Children: A Case Study on Scaling

Motivation

- Interest in the ability of RCTs to inform policy decisions
  - Banerjee et al. (2017), Niehaus-Muralidharan (2017), Al-Ubaydli et al. (2020)
- The devil is in the details (of the implementation)
  - Small variations in the protocol of the intervention often translate into substantial differences in outcomes

$\Rightarrow$ We provide a case study on scaling educational interventions

- Two RCTs + government implementation of a mentoring program
- Role in scaling of actors of child development: mentors, parents, teachers

CONAFE Schools in Mexico

- Government agency provides schooling services to poor and remote villages
  - Villages with $< 2,500$ residents
  - Small multi-grade schools (10-15 students)
  - Math and Spanish scores are 0.5-0.7 SD below the national averages
  - Less than two thirds of sixth graders enroll in secondary schooling

$\Rightarrow$ Community-based model

- Community instructors with little training and high turnover
- Parents organize local associations aimed at promoting community education

The *A*sesores *P*edagogicos *I*tinerantes (API) Program

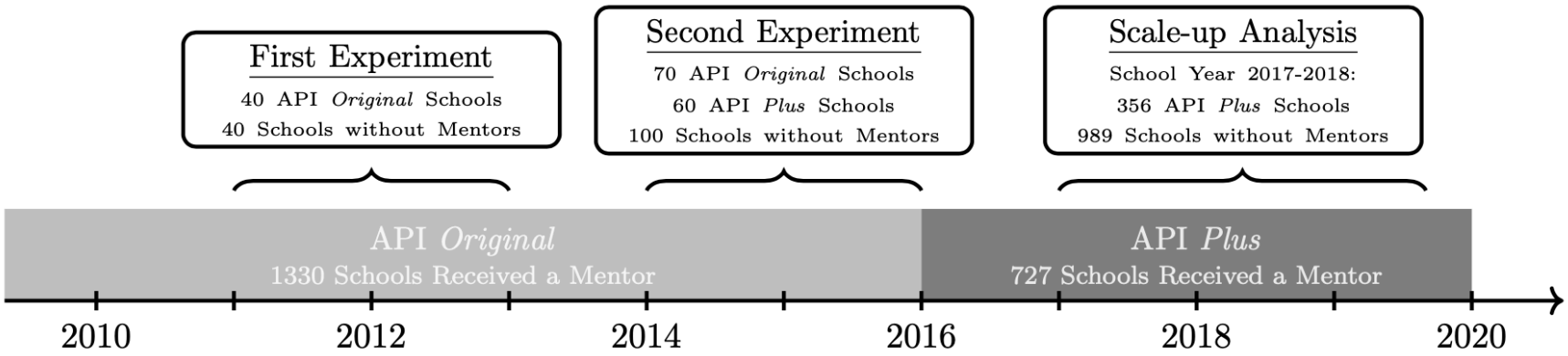
- API *Original* started in 2009 in CONAFE primary schools
  - Mentors are selected among college students/graduates
  - Initial training + periodic workshops on curricular knowledge and pedagogy
- Mentors provide educational service at the community
  1. Meetings with parents (both at home and at school)
  2. One-on-one tutoring sessions for the six weakest students in the class
  3. Pedagogical support to instructors

$\Rightarrow$ API *Plus* training includes bi-monthly workshops to improve parenting skills

- Peer-to-peer sessions (18 hours per session): sharing effective practices

Research Design

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



- Assignment of the program outside of the experiments is conditionally random
  - Criteria determining the priority of program assignments + other covariates

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

	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 at scale is in line with RCT evidence (e.g. ↑ 9.1pp vs. 12pp)
- The program's impacts endure beyond the conclusion of the 2-year intervention

Takeaways

- API *Original* is ineffective
  - Joint null across both experiments has a p-value $= 0.460$

$\Rightarrow$ API *Plus* modality has large effects on children outcomes

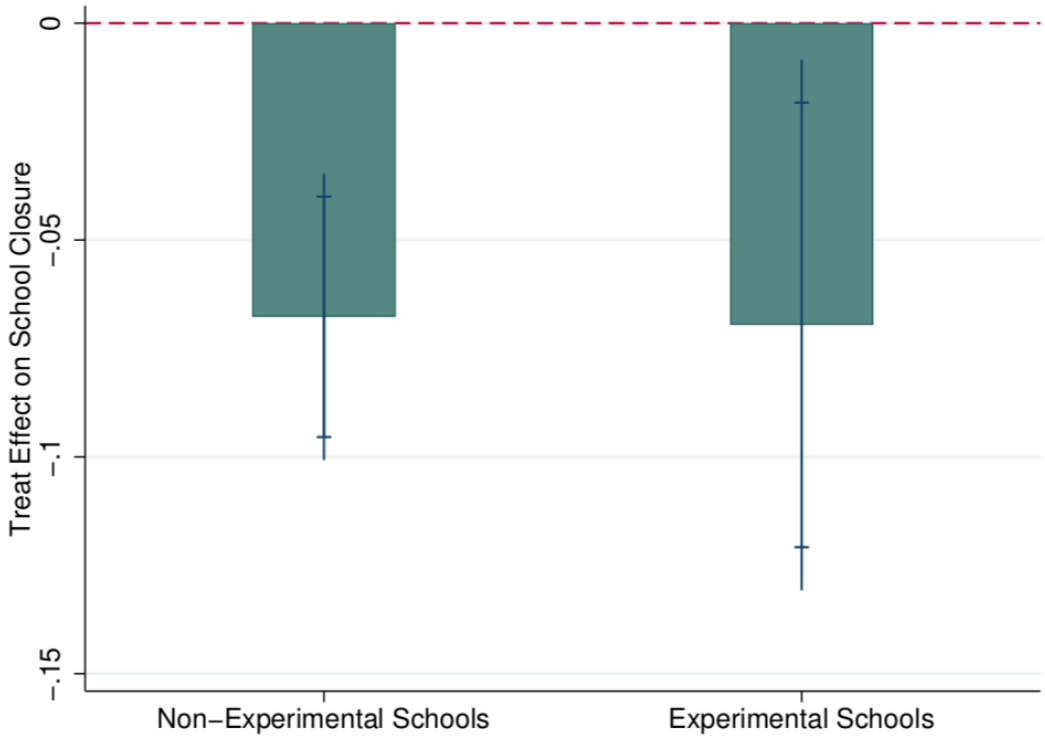
- Across different samples of schools under the same implementation regime
- Across different regimes (RCT and gov) for the same sample
- Joint null is highly significant (p-values < 0.001)

The Threat of Voltage Drop in the New Situation

- Robust elements of the evaluation
  - Experiments designed within existing infrastructure of the program at scale
  - Existing infrastructure converted into *Plus* modality
  - Randomization implemented at large unit level (school/community)
  - No site selection bias due to program targeting
- Remaining Challenges
  - Quantity and Quality of the Mentoring Service at Scale
    - No evidence of major changes in quantity or quality of mentoring practices

  $\Rightarrow$ School closures: Common disruptive event outside our experiment (11%)
    - Only two schools closed in the experimental sample of 230 schools

The Impact of the API Plus Program on School Closures



⇒ The government implementation of the *Plus* modality induces a large effect on school closures

Pathways to Scale

- Parents play an important role in the school community
- Mentors with enhanced training trigger parental engagement
  - No differences in the effectiveness of the remedial education sessions or in the pedagogical support for instructors
  - Parent/mentor and parent/child interactions as a potential mechanism

$\Rightarrow$ Community-level parental engagement may preserve "voltage" when scaling

Experimental Evidence on 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

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

- Supportive evidence for exclusion restriction
- +0.1sd of village-level parental engagement ↑ 2.2pp proba that school is open

What Drives the Success of the Program at Scale?

⇒ Active parental involvement may have helped sustain the effectiveness of the intervention at scale

- Qualitative evidence suggests parents ensure continuity in schooling activities
- Parental responses are shown to prevent schools from closing

- Persistent effects on education outcomes
  - 6.5-8.5 percent increase in enrollment in secondary school
  - 20 percent reduction of illiteracy rates

Limitations and Conclusions

- Limitations of our study
  - Silent about "vertical" aspects of scaling (e.g., government recruiting large pool of new mentors)
  - Supply-side constraints of university graduates in other contexts
- New insights on the science of scaling
  - The importance of designing RCT within existing infrastructure
  - Scaling as socially determined outcome (List et al., 2023)