Learning by Doing?
Experimental Evidence on Activity-based Instruction in India

Andreas de Barros*
Johanna Fajardo-Gonzalez†
Paul Glewwe‡
Ashwini Sankar§

Structured Abstract for the SREE Spring 2020 Conference

JEL codes: I21 Analysis of Education; I22 Educational Finance; I25 Education and Economic Development; I28 Government Policy; O12 Microeconomic Analyses of Economic Development.

Keywords: Activity-based instruction; experiential learning; India; mathematics; primary schools.

*Ph.D. Candidate, Graduate School of Education and Graduate School of Arts and Sciences, Harvard University. E-mail: adebarros@g.harvard.edu.

†Research Consultant, Poverty and Equity Global Practice, The World Bank. E-mail: jfajardog@worldbank.org.

‡Professor, Department of Applied Economics, University of Minnesota. E-mail: pglewwe@umn.edu.

§Ph.D. Candidate, Department of Applied Economics, University of Minnesota. E-mail: sanka010@umn.edu.

1 Background/Context

In a recent review, Evans and Popova (2016) conclude that pedagogical interventions and teacher training are among the most effective ways to tackle the Global Learning Crisis and to improve student learning outcomes, in the developing world. While promising, this conclusion builds on a very small evidence base. For instance, de Barros (2018) finds that, out of the 1,754 complete and ongoing trials registered at the AEA registry (501 of which study education), only 16 measure outcomes relating to pedagogy or teaching practices.

2 Purpose/Objective/Research Question

In this study, we estimate the causal effects of an innovative program in Karnataka, India, that promotes activity-based learning of mathematics through additional teaching inputs, related teacher training, and community engagement. This program is designed to enable students to learn through games, puzzles and other engaging activities, while allowing them to find creative ways to arrive at a solution—in marked contrast with the conventional chalk-and-talk method commonly used in Indian schools.

3 Setting

We observe the program in Karnataka, India, as implemented as part of a large scale-up, in partnership between the State Education Department and a local NGO, in Government schools, with public teachers, during the usual school hours.

4 Population/Participants/Subjects

Our study is being implemented in two districts in Karnataka. We purposely selected these districts to maximize the study's geographic spread and representativeness.

From all Gram Panchayats in these districts, our first step was to randomly select 49 Gram Panchayats from each district using a"probability-proportional-to size" (PPS) technique (where the selection probability reflects number of eligible schools). Within these GPs, we thereafter randomly sampled three schools each, for a total of 294 schools. Two schools were removed after baseline, reducing our sample to 292 schools.[1]

A total of 5,227 fourth-grade students were formally enrolled in the study's 292 schools. Of those, 4,026 children (77.0 percent) were present during the baseline. These 4,026 students provide the study sample.

5 Intervention/Program/Practice

The intervention combines the provision of new instructional materials, related teacher training, and community engagement to improve the mathematics abilities of primary-school students. The program was initially started with government primary schools in one block of Bangalore Rural District, in 2011. Karnataka's Government has since committed to scaling the program to all of the state's 44,000 Government primary schools, in a phased manner.

6 Research Design

We conduct a cluster-randomized trial. We conducted a stratified randomization to assign the 292 schools to be treatment or control schools. After the baseline test, within each district we used baseline test scores to create quadruplets of Gram Panchayats with similar academic performance. Thereafter, for each of these strata, two GPs were randomly selected to participate in the program, while the other two GPs remained as "controls." We then randomized all of the treatment pairs into one GP with contests and one GP without the contests.

7 Data Collection and Analysis

7.1 Data

Our primary outcome of interest is child learning after one year, in mathematics, for students in grade four (at baseline). We measure learning through three rounds (baseline, midline, endline) of tests (oral and written).

The study's secondary analyses investigate instructional behaviors, community and parental engagement, and the program's implementation fidelity. We collect these data through unannounced classroom observations. More specifically, we use a novel, standardized classroom observation instrument, developed by the World Bank,

[1]Baseline data collection revealed no students present in these two schools.

called "Teach" (Molina et al. 2018). We further complement this information with student surveys, teacher surveys, and parent surveys.

We use additional secondary data to track implementation fidelity in treatment schools, and to assess program costs.

7.2 Analysis

We estimate the intent-to-treat effect of the program on outcomes, using the following specification:

$$Y_{isg}^{t} = \alpha_g + \beta^t T_{isg} + \gamma^t Y_{isg}^{t=0} + \boldsymbol{\delta' X}_{isg}^{t=0} + \epsilon_{isg}^{t} \tag{1}$$

where Y_i is the outcome of interest for student i in school s and GP g, at time t. In our primary analysis, Y_i refers to test scores (for both tests applied in schools and GP contests.) In our secondary analyses, Y_i consists of mediating variables. The α_g parameters are randomization strata fixed effects, T_{isg} is the treatment dummy, and ϵ_{isg}^{t} is the idiosyncratic error term. To increase precision, all specifications include $Y_{isg}^{t=0}$ and $\boldsymbol{X}_{isg}^{t=0}$ as covariates. Measured at baseline ($t = 0$), $Y_{isg}^{t=0}$ refers to a student's initial outcome of interest; $\boldsymbol{X}_{isg}^{t=0}$ refers to a vector of baseline controls selected through a LASSO procedure, from student age, gender, school-level administrative data, and village-level census data (cf. Dhar et al. 2018). The coefficient of interest is β^t, which captures the program's intent-to-treat (ITT) effect, for each follow-up round t.

Additional analyses focus on (a) the causal effects of the two program components, and (b) differential effects by students' initial skill level, gender, and geographic location (i.e., district). All analyses have been pre-registered through a peer-reviewed, Registered Report.

8 Findings/Results

Our analyses of baseline data confirm that the study's randomization strategy created groups whose observable baseline characteristics are balanced, in terms of baseline performance distributions (see Figure 1), and in terms of additional observable characteristics (see Table 1 and Table 2).

[Insert Figure 1 here.]

[Insert Table 1 here.]

[Insert Table 2 here.]

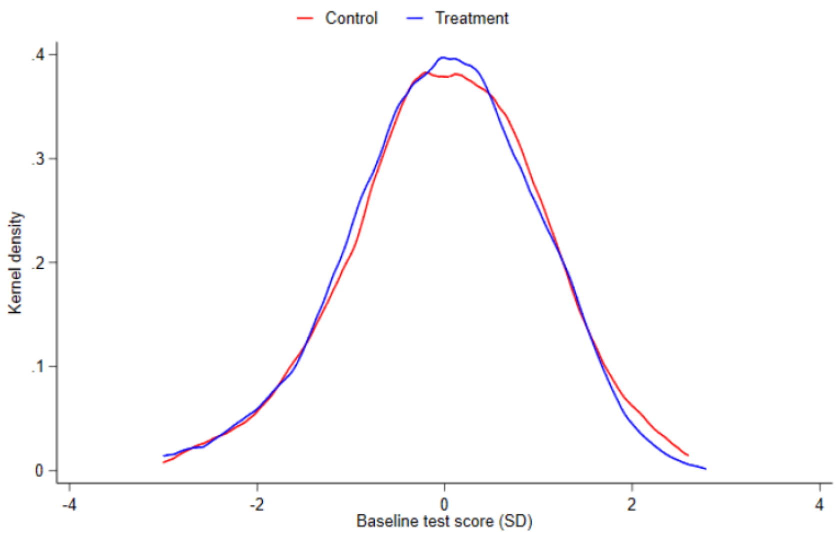
All remaining results (endline, midline, process monitoring) will be readily available by March 2020. SREE 2020 would be the first conference where we present the full set of study results.

9 Conclusion

Our paper contributes to scholarship on the determinants of learning outcomes, with particular focus on teaching quality and pedagogy. Through the use of a novel classroom observation measure, our study recognizes that prior work in this area has largely ignored program effects on teaching effort and instructional quality. We observe the program as implemented as part of a large scale-up in Government schools, with public teachers, during the usual school hours. We thus also aim to add to research on the effectiveness of public programs under government leadership, beyond smaller, tightly controlled pilots (cf. Allcott 2015; Vivalt 2019). Furthermore, our paper provides evidence on the effects of a bundled intervention that seeks "to make inputs work." We therefore also join an important avenue of emerging research that aims to answer why additional teaching inputs have often failed to produce improvements in cognitive skill (Barrera-Osorio et al. 2018; Mbiti et al. 2019).

Figures

Figure 1: Balance at baseline, overall test score (Nov. 2018)



Notes: This figure provides kernel density plots for the overall baseline test score, separately for the control and for the treatment group.

Tables

Table 1: Balance tests, student level

Variable	Control N/[Clusters]	Control Mean/[SD]	Treatment N/[Clusters]	Treatment Mean/[SD]	Difference Coef./(s.e.)
	(1)	(2)	(3)	(4)	(5)
Math Score (2PL, std.)	1948	0.031	2078	-0.029	-0.043
	[49]	[3.355]	[49]	[2.448]	(0.051)
Student ASER level	1898	2.295	2039	2.297	-0.004
	[49]	[1.397]	[49]	[1.849]	(0.032)
Percentage correct (all items)	1948	0.532	2078	0.518	-0.010
	[49]	[0.743]	[49]	[0.535]	(0.011)
Percentage correct (oral test, includes ASER as 1 'item')	1898	0.693	2039	0.694	-0.000
	[49]	[0.460]	[49]	[0.500]	(0.003)
Percentage correct (written test)	1948	0.493	2078	0.476	-0.012
	[49]	[0.827]	[49]	[0.581]	(0.014)
Percentage correct (applied, written)	1948	0.432	2078	0.414	-0.013
	[49]	[0.846]	[49]	[0.612]	(0.015)
Percentage correct (procedural, written)	1948	0.556	2078	0.539	-0.012
	[49]	[0.823]	[49]	[0.564]	(0.013)
Percentage correct (whole number operations, written)	1948	0.551	2078	0.546	-0.002
	[49]	[0.887]	[49]	[0.665]	(0.015)
Percentage correct (numbers, written)	1948	0.558	2078	0.538	-0.014
	[49]	[0.867]	[49]	[0.585]	(0.016)
Percentage correct (data, written)	1948	0.376	2078	0.353	-0.019
	[49]	[0.706]	[49]	[0.530]	(0.013)
Percentage correct (geometry, written)	1948	0.487	2078	0.469	-0.012
	[49]	[0.954]	[49]	[0.731]	(0.017)
Gender: Female	1898	0.532	2039	0.529	-0.002
	[49]	[0.538]	[49]	[0.694]	(0.017)
Student age (as of 31-Dec-18)	1901	9.140	2044	9.154	0.012
	[49]	[1.363]	[49]	[1.471]	(0.024)
F-test of joint significance (F-stat)					1.047
F-test, number of observations					3857

Notes: The value displayed for t-tests are the differences in the means across the groups.
The value displayed for F-tests are the F-statistics.
Standard deviations (SD) and number of clusters in brackets. Standard errors (s.e.) in parentheses, clustered at the Gram Panchayat level.
Randomization strata fixed effects are included in all estimation regressions.
***, ** and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: Balance tests, school level

Variable	Control		Treatment		Difference
	N/[Clusters]	Mean/[SD]	N/[Clusters]	Mean/[SD]	Coef./(s.e.)
	(1)	(2)	(3)	(4)	(5)
Percent of students appeared and passed primary exam	133	0.317	132	0.328	0.084*
	[49]	[0.457]	[49]	[0.394]	(0.050)
Female students (percentage)	146	0.497	146	0.506	-0.010
	[49]	[0.087]	[49]	[0.107]	(0.013)
Percentage OBC	146	0.645	146	0.655	-0.001
	[49]	[0.247]	[49]	[0.276]	(0.033)
Total number of teachers	146	5.192	146	5.295	0.381
	[49]	[3.126]	[49]	[3.011]	(0.312)
No of students per teacher	146	28.710	146	26.659	-1.650
	[49]	[21.980]	[49]	[16.265]	(1.571)
Female teachers (percentage)	146	0.458	146	0.414	-0.028
	[49]	[0.401]	[49]	[0.369]	(0.035)
School: Years in service	146	73.274	145	69.669	-3.158
	[49]	[27.626]	[49]	[21.380]	(3.236)
School is co-ed (vs. single-sex)	146	0.856	146	0.932	0.119**
	[49]	[0.457]	[49]	[0.312]	(0.051)
Percentage of classrooms needing minor repair	146	0.144	146	0.128	-0.003
	[49]	[0.188]	[49]	[0.187]	(0.017)
Percentage of classrooms needing major repair	146	0.126	146	0.147	0.015
	[49]	[0.219]	[49]	[0.201]	(0.019)
No. toilets / students	146	0.027	146	0.026	-0.001
	[49]	[0.030]	[49]	[0.025]	(0.001)
Boundary wall is inexistent or incomplete	146	0.534	146	0.630	0.198***
	[49]	[0.570]	[49]	[0.545]	(0.067)
School has tap water	146	0.589	146	0.603	-0.023
	[49]	[0.606]	[49]	[0.661]	(0.058)
Computers / no. of students	146	0.014	146	0.011	-0.001
	[49]	[0.027]	[49]	[0.024]	(0.002)
Received a school maintenance grant	146	0.911	146	0.884	-0.049
	[49]	[0.422]	[49]	[0.435]	(0.051)
F-test of joint significance (F-stat)					1.416
F-test, number of observations					264

Notes: The values displayed for column (5) are coefficients from regressing each variable on the treatment indicator. The value displayed for F-tests are the F-statistics.
Standard deviations (SD) and number of clusters in brackets. Standard errors (s.e.) in parentheses, clustered at the Gram Panchayat level.
***, ** and * indicate significance at the 1, 5, and 10 percent critical level.

References

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics 130*(3), 1117–1165.

Barrera-Osorio, F., S. García, C. Rodríguez, F. Sánchez, and M. Arbeláez (2018). Concentrating Efforts on Low-Performing Schools: Impact Estimates from a Quasi-Experimental Design. *Economics of Education Review 66*, 73–91.

de Barros, A. (2018). Do Students Benefit from Blended Instruction? Experimental Evidence from India.

Dhar, D., T. Jain, and S. Jayachandran (2018). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. Working Paper 25331, National Bureau of Economic Research.

Evans, D. K. and A. Popova (2016). What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. *The World Bank Research Observer 31*(2), 242–270.

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2019, August). Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania. *The Quarterly Journal of Economics 134*(3), 1627–1673.

Molina, E., S. F. Fatima, A. Ho, C. M. Hurtado, T. Wilichowksi, and A. Pushparatnam (2018). Measuring Teaching Practices at Scale: Results from the Development and Validation of the Teach Classroom Observation Tool. Working Paper 8653, The World Bank, Washington, D.C.

Vivalt, E. (2019). *How Much Can We Generalize From Impact Evaluations?* Unpublished manuscript, Australian National University, Canberra, Australia.