

CEARA TEACHER FEEDBACK PROGRAM

I. Background information

The state of Ceara, Brazil is conducting classroom observations to benchmark its teachers' use of time, materials and interactive pedagogical practices and measuring the impact that providing schools with feedback from classroom observations can have on their subsequent performance. The Secretariat's goal is to find the least-cost strategy for stimulating more interaction among teachers at the school level, that will lead to faster and cheaper diffusion of good practices and a "culture of shared practice" within schools. To generate a rigorous measure of the cost-effectiveness of this approach, the Secretariat is implementing a random assignment experiment during the 2015 school year and has requested World Bank assistance with the design and implementation of the evaluation.

A stratified representative sample of 292 secondary schools and a representative number of randomly selected teachers within each school were observed by trained observers at the end of the 2014 school year (November). Classroom practice was analyzed using the Stallings classroom snapshot, which has been extensively used by the World Bank in other parts of Brazil and other LAC countries. The Stallings method generates robust, internationally-comparable, quantitative data on four key variables: teachers' use of time, use of materials, use of interactive pedagogical practices, and ability to keep students engaged. An advantage of the Stallings method is that results can be benchmarked against good practice indicators from the US as well as other parts of Brazil and other LAC countries.

The treatment was launched at the beginning of the 2015 school year (March), in 136 schools. 156 control schools received neither information nor any supports. An endline round of classroom observations in the sample of schools and classrooms in November 2015 will capture any changes in teacher practice and correlate these with student learning results on the national and state-wide student assessments that are given at end-November. The evaluation will measure the impact of the information +coaching program on teacher practice and student learning.

II. Timeline

Baseline			Endline						
Stallings training	Baseline observations	Analysis	Toolkit development	Randomize	Treatments Implemented	Pilot of tablets	Stallings training	Endline observations	Full Analysis
Oct. 2014	Nov. 2014	Dec. 2014	Sept. 2014 - Feb. 2015	Feb. 2015	Mar-Oct. 2015	Aug. 2015	Oct. 2015	Nov. 2015	Jan-Jun. 2016

Randomization design: October 2014

Baseline: November 2014

- Stallings training course for observer team - October 20-24, 2014
note: all observers were pedagogical coordinators from treatment schools, to avoid any contamination of control schools from having someone at the school familiar with the Stallings observation method and/or the training program

- Classroom observations in 3,178 classrooms in 292 schools – Oct.27-Nov.21, 2014
- Baseline data analysis - January 2015

Bulletin development: January-February 2015

- Testing of alternative formats with focus groups of teachers and directors – February 2015
- Finalization and production for schools – March 2014

Implementation: March - October 2015

- Stallings feedback and toolkits delivered to 136 treatment schools on March 27
- School directors and pedagogical coordinators attend 3 face to face workshops
- Pedagogical coordinators and teachers interact with their ELOS coach via Skype and video uploads
- School directors and pedagogical coordinators make weekly online monitoring reports

Pilot of tablets for classroom observations: August 2014

One week test of paper-based and tablet based coding sheets by pairs of observers in same classrooms

Endline: October - November 2015

- Stallings refresher training course for observer team – October 19-23, 2015
- Final round of observations in 292 schools – Oct. 26-Nov. 13, 2015.

Full Analysis and Impact Report: January - June 2016

By March 2016:

- Analysis of school-level changes in classroom dynamics
- Analysis of school-level implementation of recommended activities (from logbooks, and teacher, student and school director surveys in a subsample of schools)
- Survey of regional supervisors

By June 2016:

Analysis of 2015 SPAECE (Ceara state student assessment), national Prova Brasil assessment (sample-based) and ENEM (high school completion test) data, whose results are available in May 2016

III. Sample Design

Randomization

Ceara state has 573 secondary schools with the capacity to participate in the in-service teacher training. A sample of 350 schools was randomly chosen from among the 573 schools, stratified by school size, geographic area and learning results (see Table 1). These schools were randomly assigned to the treatment or control group (175 in each group).

The characteristics of the treatment and control groups are balanced (see Table 2). Almost all of the variables analyzed (24 out of 25) show no significant differences between the two groups, with a significance level of 5%. For instance, the mean of the Portuguese test scores, one of the primary outcome indicators, is 257 points in treatment schools and 261 points in control schools. Only enrollment for 12th grade show a significant difference.

Baseline data

Out of the 350 schools of the randomization, 300 schools were observed in November 2014. The full number of school could not be observed because of disruptions to the school calendar in November (standardized tests and holidays), and a shortage of observers in the Fortaleza district (some coordinators did not want to participate for ideological/political reasons). After eliminating the schools with inconsistent information, the baseline data includes 292 schools: 136 in the treatment group and 156 in the control group. The observations were made in 3,176 classrooms of 10th, 11th and 12th grade. In each school, the number of classrooms observed vary from 3 to 31, depending on the school size.

Despite the reduction in the number of schools, characteristics of the treatment and control groups remained balanced (see Table 2 below). The means of key outcome indicators (test scores, pass rates and completion rates) show no significant differences between groups, with a significance level of 5%. For example, the mean of the student pass rate is 84% in the treatment group and 86% in the control group. Very importantly, the classroom dynamics indicators are also balanced (see table 3): the percentage of total class time used for instruction is 66% in the treatment group and 68% in the control group; teachers are off-task (either out of the classroom or engaged in social interaction with students or visitors) 9% of the time in the treatment and 10% of the time in control schools. In both groups, teachers use the blackboard in more than one-third of all time spent on teaching activities.

The only areas where statistically significant differences were found are school size (treatment schools are larger on average than the control schools) and the share of teachers that are female (56% in the treatment schools versus 52% in control schools). However, we do not believe that either of these differences will bias program implementation or results.

Power Analysis:

We made power calculations based on treatment and control arms of 146 schools per arm, with at least 5 teachers per school and 20 students per teacher. The minimum detectable effect sizes for the comparison are provided in table 4 for the primary outcome measure – Portuguese and math test scores – and for two models – the unconditional model and the variables conditional models (Hedges & Hedberg (2007), Bloom, Richburg-Hayes, & Black (2007), Raudenbush (2007), and Schochet (2008)). Considering an unconditional model, our analysis found a minimum detectable effect of 0.18 standard deviations for Portuguese and math test scores. Considering a set of covariates in the conditional model to explain the outcome variables, we would capture a smaller detectable effect, 0.12 sd. and 0.10 sd., respectively.

IV. Potential risks

Low intervention take-up

The treatment relies on schools' willingness and ability to identify and adopt changes in teacher behavior in response to the information and tools provided. This evaluation includes substantial emphasis on monitoring treatment take-up, since it is a crucial issue for the effectiveness of the intervention. The ELOS team is keeping records of school-level activities as well as their own log of skype conferences conducted, videos uploaded and reviewed, and feedback shared. We plan to conduct endline interviews with school principals, pedagogical coordinators and teachers to explore why some schools utilized the materials more intensively than others.

Spillovers

Since the treatment was allocated at the school level, and the sample was stratified across different municipalities state-wide, teachers in the control schools are not likely to know about or participate in the intervention. The online website for the program can only be accessed with a school code. Nevertheless, there is a small chance that some regional supervisors, who are aware of the intervention, may convey information about the program to control schools, even though they have been informed about the need to avoid this.

Attrition

Attrition could result if there are school closures in the treatment group and the control group no longer serves as a good counterfactual, or the reverse. However, there have been no school closures between 2014 and 2015 and there is a low probability of any closure during the 2015 school year.

Political risks

Even though in January 2015 a new state governor assumed office, the government's support for the study has continued. The new governor reappointed the same Secretary of Education and deputy secretary, so no key stakeholders have changed since the beginning of the study. Given Ceara's tradition of technocratic state government, we do not foresee risks of the state hiring new officials less committed to generating rigorous evaluation evidence.

V. References

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). "Using covariates to improve precision for studies that randomize schools to evaluate educational interventions". *Educational Evaluation and Policy Analysis*, 29(1), 30-59.

Hedges, L. V., & Hedberg, E. C. (2007). "Intraclass correlation values for planning group-randomized trials in education". *Educational Evaluation and Policy Analysis*, 29(1), 60-87.

Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). "Strategies for improving precision in group-randomized experiments". *Educational Evaluation and Policy Analysis*, 29(1), 5-29.

Schochet, P. Z. (2008). "Statistical power for random assignment evaluations of education programs". *Journal of Educational and Behavioral Statistics*, 33(1), 62-87.

Table 1: Sample Distribution per CREDE and learning results quartile

Test scores quartile CREDE	School Population					Random sample					Baseline data				
	1	2	3	4	Total	1	2	3	4	Total	1	2	3	4	Total
1	26	16	3	13	58	18	11	3	6	38	18	11	3	6	38
2	9	12	9	6	36	4	5	5	4	18	4	5	5	4	18
3	3	5	2	7	17	3	2	2	4	11	3	2	2	4	11
4	1	4	5	1	11	1	2	4	1	8	1	2	4	1	8
5	2	8	8	10	28	1	2	5	6	14	1	2	5	6	14
6	1	12	16	13	42	0	6	11	8	25	0	6	11	8	25
7	0	4	6	2	12	0	3	4	1	8	0	3	4	1	8
8	4	7	4	3	18	3	3	2	3	11	3	3	2	3	11
9	3	3	3	4	13	1	3	2	2	8	1	3	2	2	8
10	5	3	7	10	25	3	2	2	6	13	3	2	2	6	13
11	5	4	1	3	13	3	3	1	2	9	3	3	1	2	9
12	5	2	7	5	19	3	1	4	3	11	2	0	3	2	7
13	4	6	11	3	24	3	3	5	2	13	3	3	5	2	13
14	1	4	4	4	13	0	2	2	3	7	0	2	2	3	7
15	0	0	4	4	8	0	0	2	2	4	0	0	2	2	4
16	4	3	4	4	15	2	2	3	3	10	2	2	3	3	10
17	7	2	3	1	13	4	2	2	1	9	4	2	2	1	9
18	4	6	4	8	22	3	4	2	4	13	3	4	2	4	13
19	2	5	9	8	24	1	3	6	6	16	1	3	6	6	16
20	6	5	4	4	19	4	3	2	2	11	4	3	2	2	11
21	5	4	7	5	21	3	3	5	4	15	2	1	4	3	10
22	10	4	5	2	21	6	2	3	2	13	3	2	2	1	8
23	9	4	1	4	18	7	0	1	3	11	5	0	0	2	7
24	6	6	0	8	20	3	4	0	5	12	0	1	0	1	2
25	11	9	8	5	33	8	7	4	4	23	4	2	1	1	8
26	11	6	9	4	30	7	3	6	3	19	3	0	0	1	4
Total	144	144	144	141	573	91	81	88	90	350	73	67	75	77	292

Table 2: Balance between Treatment and Control Schools

Variables	Random sample			Baseline data		
	Treatment	Control		Treatment	Control	
	Mean	Mean	Difference	Mean	Mean	Difference
Portuguese proficiency	257.371	260.615	-3.245	256.931	261.385	-4.454*
Mathematical proficiency	267.376	272.055	-4.679	267.715	273.276	-5.562*
Average proficiency	262.377	266.334	-3.957	262.326	267.328	-5.002*
High school enrollment	641.411	586.263	55.149	676.324	575.314	101.009**
High school enrollment - regular	594.503	517.166	77.337*	628.860	497.801	131.059***
High school enrollment- vocational education	46.634	67.817	-21.183	47.110	76.077	-28.967
Enrollment 10th grade	206.994	189.617	17.377	217.471	185.583	31.887**
Enrollment 11th grade	180.183	168.211	11.971	190.265	164.878	25.387**
Enrollment 12th grade	234.463	197.223	37.240**	247.728	189.840	57.888***
Rural area	0.029	0.051	-0.023	0.037	0.058	-0.021
Pass rate	83.331	84.579	-1.248	84.459	85.574	-1.115
Failure rate	6.938	6.290	0.649	6.398	6.051	0.347
Dropout rate	9.731	9.131	0.600	9.144	8.375	0.769
Students per class	34.064	33.990	0.073	34.384	34.034	0.349
Female principals	0.520	0.514	0.006	0.485	0.519	-0.034
Experience as a principal (>10 years)	0.543	0.520	0.023	0.507	0.500	0.007
Principal with graduate degree	0.994	0.994	0.000	0.993	0.994	-0.001
Female teachers	0.551	0.517	0.034*	0.562	0.515	0.048**
Temporary teachers	0.995	0.994	0.001	0.995	0.994	0.001
Teachers' age	35.003	30.394	4.609	35.344	30.147	5.197
Experience as a teacher (>10 years)	0.816	0.814	0.002	0.819	0.812	0.007
Low salary (<2s.m.)	0.185	0.185	0.000	0.194	0.183	0.011
High salary (>5s.m.)	0.225	0.200	0.025	0.219	0.187	0.033
Mothers without EF	0.472	0.484	-0.011	0.490	0.488	0.002
Mothers with graduate degree	0.051	0.052	-0.001	0.055	0.055	0.000
Number of schools	175	175		136	156	

Table 3: Classroom dynamics in Treatment and Control Schools

	Baseline data		
	Treatment	Control	
Variables	Mean	Mean	Difference
Instructional activities	0.66	0.68	-0.018
Classroom management activities	0.25	0.23	0.022**
Off-task activities	0.09	0.10	-0.004
Reading aloud	0.043	0.04	0
Demonstration/Lecture	0.327	0.334	-0.007
Discussion/Debate/Q&A	0.097	0.099	-0.002
Practice & Drill	0.004	0.004	0
Assignment/Class work	0.123	0.132	-0.01
Copying	0.063	0.062	0.001
Verbal instruction	0.06	0.057	0.004
Discipline	0.021	0.017	0.004*
Classroom management	0.08	0.076	0.004
Classroom management alone	0.088	0.078	0.01
Social interaction	0.015	0.017	-0.002
Teacher out of the room	0.058	0.058	0
Teacher uninvolved	0.021	0.022	-0.002
No material	0.187	0.177	0.01
Textbook	0.149	0.129	0.02
Notebook	0.157	0.194	-0.037*
Blackboard	0.384	0.361	0.023
Learning aides	0.035	0.03	0.005
TIC	0.078	0.098	-0.02
Cooperative	0.009	0.01	-0.001
Number of schools	136	156	

Table 4: Minimum detectable effects for 80% power**Baseline data: 292 schools**

	Portuguese Test Scores		Math Test Scores	
	Power (κ)	Power (κ) controlling for covariates	Power (κ)	Power (κ) controlling for covariates
Number of clusters	146	146	146	146
Cluster size	100	100	100	100
Intra-cluster correlation	0.14	0.14	0.13	0.13
Significance level (α)	0.05	0.05	0.05	0.05
R^2_{level2}	-	0.55	-	0.74
Sample size	14600	14600	14600	14600
Minimum detectable effect	0.18 sd	0.12 sd	0.18 sd	0.10 sd