Enhancing Human Capital in Children:
A Case Study on Scaling
Web Appendix

Francesco Agostinelli   Ciro Avitabile   Matteo Bobba

University of Pennsylvania   The World Bank   Toulouse School of Economics

June 21, 2024

A   Data Description

A.1   Administrative Data

School census. The Ministry of Education runs a school census (*Formato 911*) at the beginning and at the end of each school cycle that covers all public schools in Mexico. The census asks the school representative about the number of students enrolled in every grade and whether they are new students or repeaters. Additional information includes the number of instructors and the number of classrooms per school. Information from the 2013 Census is used to construct the baseline school variables that are displayed in Table B-1 and in Panel A of Table B-2. School census data for the years 2015–2020 are used to track the school closures during the government implementation of both the API *Original* and *Plus* modalities, as shown in Table 7 and Figure 3.

Locality-level Population census: The National Institute of Statistics and Geography (INEGI) is in charge of compiling a population count with detailed information on socio-demographics, poverty, and education, among other information every decade. Census data are made available at the individual level for a small random sample of the population, as well as at the locality-level for the universe of localities in Mexico. We use the locality-level information collected in the census rounds of 2010 and 2020 for our analysis. In particular, we use information from the 2010 population census in Tables 1, B-1 and B-4. We leverage

information on schooling outcomes in the 2020 population census for all the localities in the state of Chiapas (including those that were part of the experimental sample), which is shown in Table 4.

Standardized test scores. Between 2007 and 2013, all Mexican students in third grades through ninth grade were required to take a standardized test, the ENLACE (*Evaluación Nacional de Logro Academico en Centros Escolares*). The test was administered by external proctors at the end of each academic year, and it assessed student knowledge in three areas: math, Spanish, and, starting in 2008, a third subject that rotated between science, ethics/civics, history, or geography. We use the school-level average of the Spanish scores in 2012 to construct the strata for the school-level randomization of the second experiment. In the first experiment, we use individual scores in each pedagogical area in 2013 as our main measures of academic achievement. The *Overall Score* displayed in Table 2 is computed using GLS-weighted score over the three scores (O'Brien, 1984). Last, we use the 2013 ENLACE scores at the school-level for the placebo tests displayed in Table B-7.

Transitions to Secondary Schools. We link the enrollment records of the sixth graders in the sample of the second experiment across the population of seventh graders in Chiapas during the following academic year. Individual transitions computed in the school year 2016–2017 (i.e., by the end of the second experiment) are reported in Table 3, while transitions computed in the school year 2017–2018 (i.e., after the first year of the government implementation of the API *Plus* modality) are reported in Figure B-2.

Other administrative records. All students in Chiapas schools, irrespective of whether they received the API program, must undergo a diagnostic test at the beginning of each school year. The test covers three subjects: math, Spanish, and natural science. The score for each subject ranges between 5 and 10. We use the individual-level average across the three subjects in the diagnostic tests at the beginning of the 2014–2015 school year to construct the within-school student rankings displayed in Figure B-3 and Table B-14, which proxy for the individual eligibility for the one-on-one remedial education sessions.

We use student-level longitudinal information for the population of primary schools to construct various measures of school-level changes in student composition reported in Table B-3: whether the student must repeat a grade in school year 2015–2016, attrition from the school system in Chiapas between the school years 2014–2015 and 2015–2016, and whether in 2015-2016 the student attends the same school as in 2014–2015.

A.2 Survey Data

Measures of Children's Achievement. We use the Early Grade Reading Assessment (reading score) and the Early Grade Math Assessment (math score) as our main measures of children's cognitive achievement. Those are individually administered student assessments that have been conducted in more than 40 countries and in a variety of languages (Dubeck and Gove, 2015; Platas et al., 2016). While these instruments are typically applied to students in first, second, or third grade, we administer them to third through six grade students to account for the large learning gaps of the children in our sample. The school-average standardized scores in math and Spanish as measured in the school year prior to the introduction of the second experiment are, respectively, 0.5 and 0.7 standard deviations below the national averages.[1] The reading scores reported in Tables 3 and B-14 are given by the latent factor of an exploratory factor analysis of the following eight domains: 1) letter name, 2) initial name, 3) initial sound, 4) word recognition, 5) word reading, 6) reading comprehension, 7) listening, 8) dictation. The math scores reported in Tables 3 and B-14 are given by the latent factor of an exploratory analysis of the following seven domains: 1) number identification, 2) number discrimination, 3) missing number, 4) addition, 5) subtraction, 6) problem solving, 7) shape recognition. An orthogonal rotation is applied before standardizing each factor with respect to the mean and the standard deviation in the control group. The individual components of the math and reading scores are reported in Table B-8.

To measure the impact of the intervention on socio-emotional skills, we consider a collection of thirty-two behavioral issues as reported by a caregiver, which resembles the questionnaire in the Children section of the National Longitudinal Study of Youth (CNLSY-79), such as antisocial behavior, anxiety/depression, headstrongness, hyperactivity and peer conflicts (for details, see Appendix A.2). The resulting behavioral problem index is re-scaled in such a way that higher values are associated with fewer behavioral issues (socio-emotional score). The survey also contains a module on instructors' characteristics as well as pedagogical practices collected through an adapted version of the Stallings Classroom Snapshot (Bruns and Luque, 2015), a module on parental attitudes and investment toward children's education, as well as information about the mentors' activities in the communities, among others. To better

---

[1] Only 5 percent of the children in our sample score at the maximum of the scale in two or more subdomains of the reading score (out of eight subdomains) and in three or more subdomains of the math score (out of a total of seven subdomains). Unlike the first experiment, we cannot leverage the national standardized test scores for the second experiment since the test ceased to be universal during the period of interest (after 2014).

interpret our results, we standardize most of the survey-based outcome variables using the mean and the standard deviation observed in the control group. The socio-emotional scores reported in Tables 3 and B-14 are the sum of the following thirty-two items on how often the child displays a given emotion/behavior: 1) has serendipitous mood changes, 2) feels or complains that nobody loves him/her, 3) is tense or nervous, 4) lies or cheats, 5) is scared or anxious, 6) talks and argues too much, 7) has difficulty focusing on a specific activity for an extended amount of time, 8) gets easily confused, 9) has his/her head is in the clouds, 10) threatens or is mean with other children, 11) tends to challenge parental authority, 12) does not feel guilty after a bad deed, 13) does not get along with other children, 14) is impulsive or acts "fast" without thinking, 15) has inferiority issues, 16) has no friends, 17) has difficulty letting go of certain thoughts, 18) is hyper active, 19) has a bad temper or is irascible, 20) easily loses his/her temper, 21) feels unhappy, sad, or depressed, 22) is shy, does not socialize with others, 23) breaks objects on purpose, 24) is too attached to adults, 25) cries too much, 26) demands a lot of attention, 27) is too much dependent on others, 28) is afraid of other people's judgment, 29) tends to be in bad company; 30) reserved, keeps things for himself/herself, 31) worries about everything, 32) misbehaves at school and does not respect the instructor.

The *Overall Score* of students' achievement displayed in Table 3 is computed using GLS-weighted averages over the two cognitive measures and the socio-emotional score.

Parenting Practices. The household survey collects information on parents' behavior and investment in their children's education. The same information was collected during the mid-line survey of the first experiment. The parental engagement outcomes reported in Table 5 are computed using GLS-weighted averages over different indicators of parental behavior. For *Engage at School*: whether or not parents (i) volunteer at the school, (ii) donate money to the school, (iii) donate in kind to the school, and (iv) offer food to the instructor. For *Manage School Resources*: whether or not parents (i) directly manage the school budget, (ii) propose some materials to the school, (iii) decide to use some materials for the school, and (iv) decide on how to allocate money for some school activities, and (v) define the pedagogical targets of the school. For *Engage with Child*: whether (i) parents help with their child's homework, (ii) meet with the instructor, (iii) expect their child to complete secondary education or more, and (iv) children participate in other academically-related activities outside the school hours. The *Engagement Index* is the same GLS-weighted average over each of the individual components described above, which are reported in Table B-13.

Mentor Characteristics. As part of the data collection activities, we have collected basic socio-demographic information on the mentors who served in the schools of the second experiment. Those are reported in Panel C of Table B-2 and in the second column Table B-10. For the other schools in Chiapas that were not part of the experimental sample, we rely on administrative rosters about mentors' characteristics from the program. Those are reported in the first column of Table B-10.

Parent-Mentor Interactions. The household module collects several questions on both the quantity and the quality of parents' interactions with the mentors for those households that were assigned to either the API *Original* group or the API *Plus* group. This information is used to construct the four variables reported in Panel A of Table 6. Basic information on both the household module respondent and household characteristics is reported in Panel B of Table B-2.

Parenting Styles. The mentors' questionnaire included a battery of questions on the specific competencies they promote during their interactions with parents. The indicator variables for each competency are used as outcomes variables in Panel B of Table 6.

Teaching Practices. Local instructors were asked standard questions on their socio-demographic characteristics, education and experience. Those are reported in Panel A of Table B-2. We measure time use and different learning activities of community instructors as well as their ability to keep students engaged using an adapted version of Stallings classroom snapshot, which is a rubric for timed observations that has been used previously in Mexico (Bruns and Luque, 2015). An observer scores the instructor's effective use of 15 different activities over the course of a full one-hour lesson, with snapshots every three minutes. Each activity was scored between 1 and 4. In every snapshot, the external observer reports whether the instructor is present in the classroom. Given the nature of the API intervention and the multi-grade context, the tool was adapted to capture the instructor's ability to use materials and keep the rhythm of the class.

The information included in this survey module is used to construct GLS-weighted averages over the different types of teacher behavior, which are displayed in Table B-15. *Learning Activities* is the sum of the amount of time children spend on (i) reading aloud alone, (ii) reading aloud in a group, (iii) questions and answers, (iv) memorizing, (vi) individual homework, and (viii) verbal tasks. *Engage with Students* is the sum of the amount of time the instructor spends on (i) elaborating on a given concept, (ii) students were not involved, and (iii) keeping discipline. *Manage Time* is the amount of time the instructor spends (i)

out of the classroom, (ii) effectively administering some tasks in the classroom, (iii) whether or not the instructor complies with the start and end time of each classroom, (iv) whether or not the instructor keeps the rhythm of the class as well as of the individual students according to their age and their mother-tongue, and (v) whether or not the students were grouped according to their respective academic levels. *Use of Material* is the sum of four indicator variables: (i) whether the instructor uses any book to explain a given topic, (ii) whether the instructor uses any material from the community to explain a given topic, (iii) whether drawings and other students' artworks are displayed in the classroom, and (iv) whether charts and maps are displayed in the classroom. The *Overall Index* is the same GLS-weighted average of the individual components of teacher behavior described above.

Quantity and Quality of Mentoring Services. Local instructors were asked about mentors' practices and activities within the local communities at two specific points in time: during the end-line survey of the experiment (Spring 2016) and in an additional follow-up survey module conducted in the fall of 2018 among the schools that were previously involved in the second experiment. The end-line survey was conducted in 57 out of a total of 58 schools that received the API *Plus* during the experiment. The follow-up survey was conducted in 93 out of a total of 103 schools that received the API *Plus* program during the government's program implementation. We obtained information about mentors from the responses collected by local instructors for 56 schools in the end-line survey and 58 schools in the follow-up survey. The corresponding measures are presented in Figure 2 and Table B-11.

A.3   In-Depth Interviews

In the spring of 2022 we implemented a series of semi-structured phone interviews with a small sample of local instructors and mentors who participated in the program. In total, we were able to locate and contact 104 local instructors and 68 mentors. Of those, 12 instructors and 16 mentors agreed to complete the phone interview. More than half of the survey respondents continued working as mentors after the 2016 government implementation of the *Plus* modality. The characteristics of the survey respondents in comparison with the overall sample are shown in Tables A-1 and A-2.

The survey contains a series of open questions related to the experiences of the mentors/local instructors with the parents in the communities. Below, we report the original quotes in Spanish that we refer to in the main body of the paper (authors' translation from Spanish).

In particular, these quotes from the mentors about the peer-to-peer sessions of the training are reported in Section 4.2:

> "*Fue un momento de la capacitación en donde me dijeron que debía adaptarme al contexto de su centro del trabajo, de comprender las necesidades y de entender situaciones que se vivían en la misma comunidad, para poder dialogar con los padres y atender a los niños sin afectar o modificar lo que ellos conciben como su medio.*"

> "*Recomendaban hacer las visitas domiciliarias con frecuencia y ayudarle en algo a los papás o salían con ellos a visitas y les daba más confianza.*"

> "*[Las sesiones de orientacion me permitieron] escuchar las diferentes estrategias que ellos tenían para poder probarlas e implementarlas.*"

These quotes from the local instructors about the role of parents in the day-by-day routine of the school are reported in Section 4.3.

> "*La gestión dela escuela y se le hicieron mejoras de cercado, pintaron la escuela arreglaron los baños y se compraron materiales.*"

> "*Eran participativos, estaban pendientes del bienestar de la escuela por ejemplo la construcción, de materiales e incluso de los desayunos y alimentación del instructor.*"

> "*Los padres apoyaban en el seguimiento al bloc de tareas y trabajaban en equipo cuando los API que no podían estar presentes por apoyar a otra comunidad, los mantenían al corriente o, incluso un poco más avanzados, por lo que cuando los APIs regresaban podían dar continuidad a sus clases sin ningún atraso.*"

Table A-1: Characteristics of Mentors—Sample vs Phone Survey

	Original Sample	2022 Survey	Difference
Age	28.443	27.556	0.888
	(3.260)	(3.941)	(1.150)
Male	0.585	0.778	-0.193
	(0.495)	(0.441)	(0.171)
High School Completed	0.868	1.000	-0.132
	(0.340)	(0.000)	(0.114)
Training Weeks	2.858	2.667	0.192
	(2.035)	(1.871)	(0.703)
Experience as Api	21.274	13.444	7.829
	(10.058)	(6.803)	(3.425)
Previously Local Instructor	0.840	0.778	0.062
	(0.369)	(0.441)	(0.130)
Previously Education Assistant	0.085	0.000	0.085
	(0.280)	(0.000)	(0.094)
Days Spent in the Community	13.528	13.556	-0.027
	(5.331)	(4.876)	(1.840)
Students Lagging Behind	5.698	5.889	-0.191
	(1.657)	(3.018)	(0.621)

Notes: This table reports means and standard deviations for the characteristics of the mentors in the main sample of the analysis and those of the mentors who participated in the in-depth phone interviews (2022). The differences reported in the last column of the table are based on OLS estimates of the regression models that control for stratification dummies. Standard errors of the mean differences for the student characteristics are reported in parentheses in the last column and they are clustered at school level. For detailed descriptions of the survey variables used in this table, see Appendix A.2.

Table A-2: Characteristics of Local Instructors—Sample vs. Phone Survey

	Original Sample	2022 Survey	Difference
Age	21.284	21.157	0.127
	(2.585)	(2.034)	(0.702)
Male	0.560	0.786	-0.226
	(0.497)	(0.426)	(0.135)
Lower than Upper Second	0.062	0.071	-0.010
	(0.241)	(0.267)	(0.066)
Upper Second Complete	0.800	0.643	0.157
	(0.401)	(0.497)	(0.111)
Above Upper Second	0.138	0.286	-0.148
	(0.346)	(0.469)	(0.097)
Experience in Months	13.545	13.429	0.117
	(9.408)	(9.362)	(2.577)
Training Weeks at Baseline	4.768	5.500	-0.732
	(4.114)	(5.019)	(1.140)
Time spent in the School	9.509	9.071	0.438
	(4.220)	(3.269)	(1.146)
Sleeps in the Community	0.651	0.857	-0.206
	(0.478)	(0.363)	(0.130)
Nights spent in the Community	3.204	3.071	0.132
	(2.065)	(2.093)	(0.566)

Notes: This table reports means and standard deviations for the characteristics of the mentors in the main sample of the analysis and those of the mentors who participated in the in-depth phone interviews (2022). The differences reported in the last column of the table are based on OLS estimates of the regression models that control for stratification dummies. Standard errors of the mean differences for the student characteristics are reported in parentheses in the last column and they are clustered at the school level. For detailed descriptions of the survey variables used in this table, see Appendix A.2.
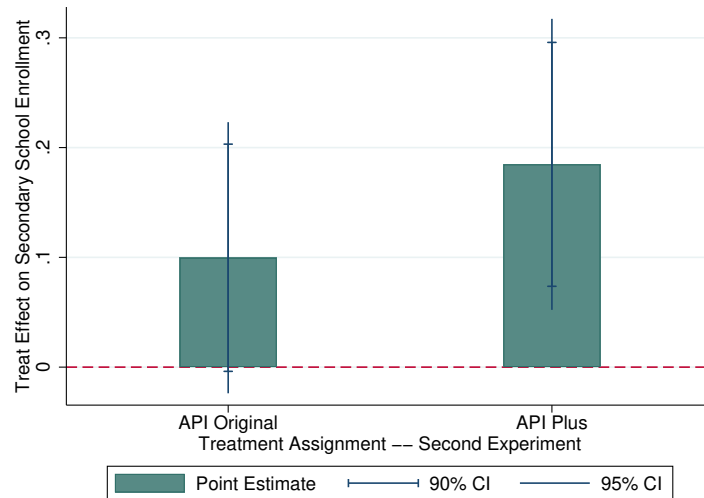
B   Additional Figures and Tables

Figure B-1: Treatment Effects on Secondary School Enrollment During the Transition Between the Second Experiment and the Government Implementation

Notes: The bars depicted in this figure show the OLS estimates of the treatment assignments in the second experiment on the probability of enrolling in seventh grade in the year after the end of the second experiment (2017). The vertical lines overlaid on the bars represent asymptotic confidence intervals at the 90 percent and the 95 percent confidence levels. Confidence intervals are based on asymptotic inference. The sample includes 207 schools of the 224 that were part of the experiment. Beyond a school that permanently closed, the sample attrition is caused by schools not having sixth graders during that school year. Attrition is balanced among schools that were part of the two treatment arms (*p*-values = 0.914, and 0.768).

Figure B-2: The Cumulative Effect of API *Plus* in the Experimental Sample of Schools
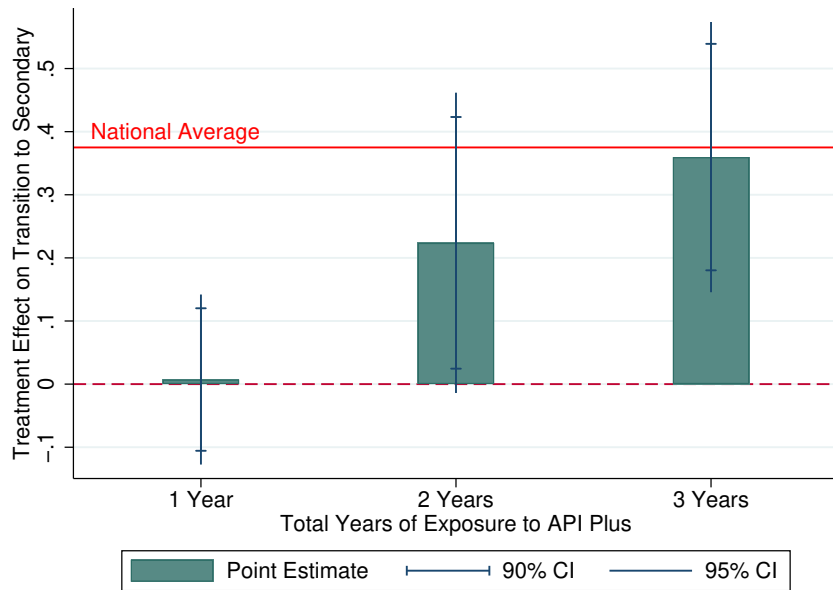
Notes: This figure shows OLS estimates of the years of exposure to the mentoring program on the probability of enrolling in seventh grade during the transition from the second experiment to the government implementation of the API *Plus* modality. Vertical lines overlaid on each bar display the 95 percent and 90 percent confidence intervals, respectively. Confidence intervals are based on asymptotic inference. The sample includes 207 schools of the 224 that were part of the second experiment. Beyond a school that permanently closed, the sample attrition is caused by schools not having sixth graders during that school year. Attrition is balanced with respect to the indicator variables for the years of exposure to API *Plus* (*p*-value[1 year]=0.467, *p*-value[2 years]=0.812, and *p*-value[3 years]=0.568, the reference category is zero years of exposure).

Figure B-3: Probability of Being in Remedial Sessions by Inverted Achievement Rank
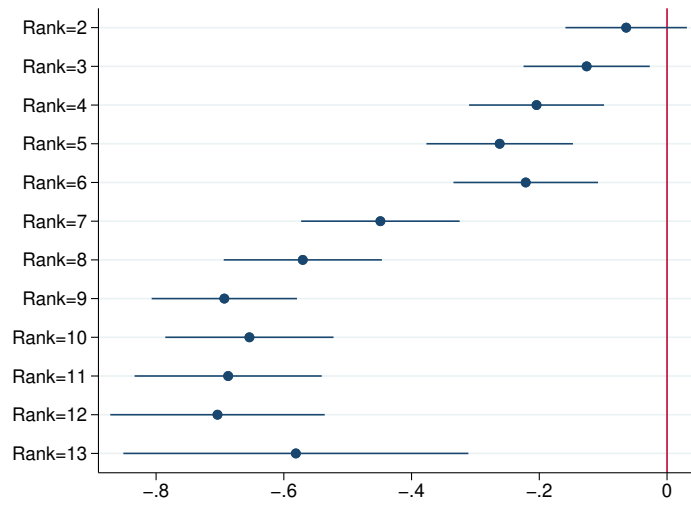
Notes: The dots in this figure are estimated marginal effects from Probit regression models of indicator variables for the inverted within-school student rank based on the average score on the diagnostic tests in math, Spanish, and natural science on the probability of participating in the one-on-one remedial education sessions with the mentors. The indicator variable for whether the student is ranked first (i.e., the worst-performing student in the class) is the omitted category. The horizontal lines around each dot represent 90 percent confidence intervals. Confidence intervals are based on asymptotic inference.

Table B-1: Baseline Characteristics and Covariate Balance – First Experiment

	API Original		Control		Diff
	Mean	Std. Dev.	Mean	Std. Dev.	*p*-value
Panel A: Schools in Mid-Line 2012 Survey of Parents					
Average Test Score (Spanish)	401.971	38.973	399.036	28.974	0.703
Average Test Score (Math)	377.916	43.159	388.422	51.038	0.351
Number of Students	15.917	8.334	14.917	7.987	0.597
Number of Teachers	1.389	0.549	1.417	0.604	0.827
Share Over-aged Students	2.134	7.225	1.961	4.094	0.900
Total Population	217.054	597.061	234.778	506.694	0.888
Labor Force Participation	0.286	0.064	0.276	0.069	0.553
Water Network (Y/N)	0.027	0.164	0.056	0.232	0.547
Sewer System (Y/N)	0.027	0.164	0.028	0.167	0.990
Rate of Illiteracy	0.321	0.170	0.333	0.173	0.745
Garbage Collection (Y/N)	0.027	0.164	0.056	0.232	0.551
Number of Schools/Localities	37		36		
Panel B: Schools with Individual Test Score 2013 Data					
Average Test Score (Spanish)	401.869	40.034	399.206	29.378	0.748
Average Test Score (Math)	377.168	44.284	390.561	50.120	0.242
Number of Students	15.971	8.449	14.743	8.034	0.527
Number of Teachers	1.400	0.553	1.400	0.604	1.000
Share Over-aged Students	2.195	7.321	2.017	4.140	0.900
Total Population	225.857	612.996	227.543	512.201	0.990
Labor Force Participation	0.287	0.065	0.278	0.069	0.579
Water Network (Y/N)	0.029	0.169	0.057	0.236	0.568
Sewer System (Y/N)	0.029	0.169	0.029	0.169	1.000
Rate of Illiteracy	0.327	0.165	0.335	0.175	0.823
Garbage Collection (Y/N)	0.029	0.169	0.057	0.236	0.566
Number of Schools/Localities	35		35		

Notes: This table shows means and standard deviations for community and school characteristics collected in the population census (2010) and the school census (2010). See Appendix A.1 for more details on these data sources. The fifth column reports the associated *p*-values of the differences in means between the treatment and the control group.

Table B-2: Baseline Characteristics and Covariate Balance – Second Experiment

Sample Statistic	Control Mean (SD)	API Original Mean (SD)	API Plus Mean (SD)	Original-Control Mean Difference (*p*-value)	Plus-Control Mean Difference (*p*-value)
Panel A: School and Teacher Characteristics					
Average Test Score (Spanish)	429.389	432.326	430.573	0.846	0.743
	(60.477)	(67.579)	(67.463)	(0.738)	(0.792)
Average Test Score (Math)	453.090	455.820	451.627	0.156	-2.057
	(78.436)	(84.546)	(82.461)	(0.978)	(0.778)
Average Test Score (Science)	438.349	441.259	442.856	1.435	3.866
	(50.264)	(49.323)	(50.492)	(0.735)	(0.390)
Number of Teachers	1.224	1.309	1.207	0.086	-0.016
	(0.419)	(0.465)	(0.409)	(0.213)	(0.820)
Number of Students	15.296	15.441	14.379	0.161	-0.953
	(5.819)	(5.655)	(5.824)	(0.857)	(0.320)
Teacher with Secundary Education	0.763	0.794	0.833	0.031	0.072
	(0.389)	(0.398)	(0.358)	(0.628)	(0.241)
Years of Experience as Teacher	0.737	0.706	0.693	-0.034	-0.042
	(0.872)	(0.802)	(1.085)	(0.802)	(0.797)
Months of Teacher Working in the School	9.531	9.309	9.281	-0.229	-0.249
	(3.947)	(4.925)	(3.266)	(0.751)	(0.676)
Observations	98	68	58	166	156
Panel B: Child and Household Characteristics					
Age in Months at Baseline (September 2014)	104.993	104.289	105.539	-0.818	0.647
	(16.384)	(17.532)	(14.924)	(0.485)	(0.605)
Male (Y/N)	0.532	0.519	0.543	-0.011	0.013
	(0.500)	(0.500)	(0.499)	(0.734)	(0.772)
Indigenous Language (Y/N)	0.302	0.307	0.461	0.012	0.155
	(0.460)	(0.462)	(0.499)	(0.855)	(0.032)
Scholarship (Y/N)	0.746	0.733	0.747	-0.013	0.005
	(0.436)	(0.443)	(0.435)	(0.763)	(0.903)
Parent Can Read	0.715	0.686	0.734	-0.030	0.023
	(0.452)	(0.465)	(0.443)	(0.465)	(0.590)
Parent with Less than Primary	0.614	0.587	0.584	-0.027	-0.029
	(0.487)	(0.493)	(0.494)	(0.526)	(0.483)
Household Receives Oportunidades CCT	0.812	0.807	0.829	-0.003	0.016
	(0.391)	(0.395)	(0.377)	(0.929)	(0.614)
Observations	453	322	269	775	722
Panel C: Mentor Characteristics					
Age in Years		28.386	28.400		0.242
		(3.678)	(3.057)		(0.705)
Male		0.579	0.620		0.051
		(0.498)	(0.490)		(0.597)
High Edu Complete		0.877	0.880		0.006
		(0.331)	(0.328)		(0.926)
Months of Experience as Mentor		22.298	20.040		-2.218
		(10.997)	(8.755)		(0.260)
Observations		57	50		107

Notes: The first three columns of the table report mean and standard deviations in parentheses for various characteristics collected before the assignment of the API program in the evaluation sample. The school variables in Panel A are computed from the 2013 national standardized tests and from the 2013 school census. The other characteristics reported in Panels B-D are collected in the survey data. The differences reported in the last two columns of the table are based on OLS estimates of the regression models that control for stratification dummies. *p*-values for the null hypothesis of equal mean differences are reported in parentheses in the last two columns. See Appendix A for more details on the data sources.

Table B-3: Treatment Assignment and School-Level Student Composition

	Repeat	Attrition	Outside CONAFE in $t-1$	Same school in $t-1$
API Original	-0.011	-0.018	-0.002	0.019
	[0.132]	[0.324]	[0.898]	[0.291]
API Plus	-0.010	-0.006	-0.003	0.011
	[0.165]	[0.760]	[0.825]	[0.573]
p-value H0: API Standard=API Plus	[0.834]	[0.491]	[0.911]	[0.620]
Observations	1019	1019	1019	1019
Clusters	224	224	224	224

Notes: This table shows the estimates of the two API modalities on various measures of school-level changes in student composition. The number of observations drops from 1045 to 1019 due to incomplete school identifiers (CURP) for 26 students. Asymptotic p-values reported in brackets are clustered at school level. For a detailed descriptions of the schooling records used in this table, see Appendix A.1.

Table B-4: Differences Between Overall Samples and Matched-census Samples

	All Chiapas			Second Experiment		
	All Sample	Census Sample	Mean Difference	All Sample	Census Sample	Mean Difference
	Mean (SD)	Mean (SD)	(p-value)	Mean (SD)	Mean (SD)	(p-value)
			Panel A: School Characteristics			
Average test score (Spanish)	424.503	422.903	1.600	431.340	433.855	-2.515
	(56.466)	(54.786)	(0.522)	(60.810)	(63.370)	(0.705)
Average test score (Math)	414.921	413.736	1.184	421.333	424.043	-2.710
	(75.300)	(74.699)	(0.725)	(80.895)	(84.848)	(0.760)
Number of students	14.049	13.974	0.075	15.009	15.158	-0.149
	(8.468)	(8.865)	(0.834)	(6.053)	(5.794)	(0.799)
Number of Teachers	1.231	1.240	-0.008	1.217	1.217	-0.000
	(0.467)	(0.480)	(0.671)	(0.413)	(0.414)	(1.000)
Share Over-aged Students	0.349	0.348	0.001	0.324	0.290	0.034
	(0.797)	(0.818)	(0.971)	(0.659)	(0.615)	(0.589)
			Panel B: Locality Characteristics			
Total Population	118.758	121.170	-2.412	121.389	158.276	-36.887
	(221.648)	(208.666)	(0.775)	(240.562)	(337.620)	(0.219)
Share of High-Poverty Villages	0.490	0.489	0.001	0.473	0.453	0.020
	(0.500)	(0.500)	(0.945)	(0.500)	(0.499)	(0.702)
Incidence of Social Conflict (Y/N)	0.190	0.204	-0.014	0.187	0.201	-0.014
	(0.392)	(0.403)	(0.355)	(0.391)	(0.402)	(0.719)
Share of Illiterate Adults	0.313	0.315	-0.002	0.295	0.292	0.003
	(0.160)	(0.159)	(0.703)	(0.153)	(0.150)	(0.860)
Share of Adults in the Labor Force	0.297	0.296	0.002	0.303	0.301	0.002
	(0.076)	(0.077)	(0.575)	(0.070)	(0.067)	(0.765)
Locality Access without Road	0.216	0.224	-0.008	0.179	0.149	0.029
	(0.411)	(0.417)	(0.609)	(0.384)	(0.357)	(0.426)
Water Network (Y/N)	0.028	0.028	0.000	0.022	0.038	-0.016
	(0.164)	(0.164)	(0.998)	(0.146)	(0.192)	(0.341)
Sewage System (Y/N)	0.011	0.012	-0.001	0.009	0.016	-0.008
	(0.105)	(0.109)	(0.830)	(0.093)	(0.127)	(0.497)
Garbage Collection (Y/N)	0.022	0.023	-0.002	0.022	0.027	-0.005
	(0.146)	(0.151)	(0.784)	(0.146)	(0.163)	(0.724)
Observations	1,523	1,161	3,046	230	184	414

Notes: Means and standard deviations in parentheses for various characteristics collected before the introduction of the API program. The last column shows asymptotic p-values for mean differences between the overall population and the experimental sample. Panel A shows community-level characteristics from the population census (2010), whereas Panel B displays school-level variables from the school census (2010). See Appendix A.1 for more details on the data sources.

Table B-5: Heterogeneity in the Impact of the Program by Eligibility Criteria

	Achievement Index	Enrolled Secondary
API Original	0.095	0.134
	[0.382]	[0.209]
API Original× Poverty	0.069	-0.098
	[0.712]	[0.441]
API Original× Welfare	0.097	-0.035
	[0.464]	[0.754]
API Original× Achievement Score	-0.036	0.003
	[0.689]	[0.964]
API Plus	0.329	0.276
	[0.039]	[0.022]
API Plus× Poverty	0.080	0.014
	[0.682]	[0.909]
API Plus× Welfare	-0.001	-0.220
	[0.997]	[0.124]
API Plus× Achievement Score	-0.058	-0.018
	[0.569]	[0.797]
API Original(Poverty)=API Original(Welfare)=API Original(Score)	[0.732]	[0.753]
API Plus(Poverty)=API Plus(Welfare)=API Plus(Score)	[0.774]	[0.450]
Observations	1045	468
Clusters	224	182

Notes: This table shows OLS estimates and the associated asymptotic p-values (in brackets) on student outcomes measured after two academic years of exposure to the API program under the second experiment designed and implemented by the authors in collaboration with the government. For a detailed descriptions of the test score index used in this table, see Appendix A.2. The dependent variables in the first column is standardized with respect to its mean and the standard deviation in the control group. The dependent variable in the second column is computed from administrative school records (see Appendix A.1). All p-values account for clustering at the school level.

Table B-6: Program Assignment at Scale (2017-2018) and Observable Characteristics

Difference	
	OLS Estimates [p-value]
Panel A: School Characteristics	
Average test score (Spanish)	-3.635
	[0.395]
Average test score (Math)	-7.707
	[0.184]
Number of students	0.644
	[0.264]
Number of Teachers	0.038
	[0.263]
Share Over-aged Students	0.020
	[0.690]
Panel B: Locality Characteristics	
Total Population	8.847
	[0.515]
Share of High-Poverty Villages	-0.035
	[0.296]
Incidence of Social Conflict (Y/N)	0.022
	[0.100]
Share of Illiterate Adults	-0.015
	[0.121]
Share of Adults in the Labor Force	-0.004
	[0.472]
Locality Access without Road	-0.061
	[0.026]
Water Network (Y/N)	0.007
	[0.538]
Sewage System (Y/N)	0.012
	[0.137]
Garbage Collection (Y/N)	-0.002
	[0.841]
F-Statistic for Joint Hypothesis of no Differences	0.76
	[0.733]
Observations	1,345

Notes: This table shows OLS estimates and asymptotic *p*-values (in brackets) for the indicator of the API *Plus* assignment during the 2017-2018 school cycle after controlling for the assignment criteria (Section 2.1), an indicator variable for prior exposure to the API Original modality, and the number of hostile event related to land property, religion, elections, crime, or drug addiction as reported at the locality level in the population census (2010). Panel A shows community-level characteristics from the population census (2010), whereas Panel B displays school-level variables from the school census (2010). See Appendix A.1 for more details on the data sources.

Table B-7: Placebo Test for API Plus Assignment During Program Scale-up

	Spanish		Math		Science	
API Plus	-0.104	0.003	-0.093	-0.001	-0.062	0.027
	[0.062]	[0.954]	[0.099]	[0.989]	[0.268]	[0.621]
	{0.057}	{0.953}	{0.101}	{0.990}	{0.268}	{0.633}
	(0.121)	(0.993)	(0.151)	(0.993)	(0.270)	(0.878)
Observations	1183	1183	1183	1183	1183	1183

Notes: This table shows OLS estimates and the associated p-values of the assignment API *Plus* in the fall of 2017. For detailed descriptions of the 2013 school-average test scores used in this table as outcome variables, see Appendix A.1. Control variables include indicator functions for the four criteria used to determine the differential priority across eligible schools to receive the mentors (see Section 2.1) as well as an indicator function for prior exposure to the mentoring program and the number of hostile event related to land property, religion, elections, crime, or drug addiction as reported at the locality level in the population census (2010). p-values reported in brackets refer to the conventional asymptotic standard errors. p-values reported in braces are computed using randomization inference (randomization-t). p-values reported in parentheses are adjusted for testing the null impact of API Plus across the two specifications considered (without and with controls) through the step-wise procedure described in Romano and Wolf (2005a,b, 2016).

Table B-8: Average Program Impacts by Subdomains of the Reading and the Math Scores

	Letter Name	Initial Name	Initial Sound	Word Recogn.	Word Reading	Read Comprehen.	Listening	Dictation
Panel A: Share of Correct Reading Answers by Subdomain								
API Original	0.103	0.006	0.122	0.129	0.075	0.118	-0.004	0.129
	[0.232]	[0.941]	[0.156]	[0.091]	[0.300]	[0.107]	[0.963]	[0.120]
	{0.285}	{0.949}	{0.194}	{0.124}	{0.341}	{0.138}	{0.968}	{0.173}
	(0.449)	(0.996)	(0.365)	(0.255)	(0.510)	(0.290)	(0.996)	(0.314)
API Plus	0.240	-0.019	0.042	0.318	0.197	0.321	0.123	0.378
	[0.005]	[0.816]	[0.565]	[0.000]	[0.014]	[0.000]	[0.145]	[0.000]
	{0.010}	{0.824}	{0.584}	{0.000}	{0.026}	{0.001}	{0.185}	{0.000}
	(0.005)	(0.789)	(0.728)	(0.000)	(0.021)	(0.000)	(0.226)	(0.000)
API Original = API Plus	[0.180]	[0.771]	[0.343]	[0.039]	[0.183]	[0.023]	[0.094]	[0.005]
	{0.174}	{0.799}	{0.479}	{0.062}	{0.229}	{0.059}	{0.220}	{0.003}
	(0.328)	(0.727)	(0.421)	(0.077)	(0.328)	(0.045)	(0.194)	(0.010)
Observations	1044	1044	1044	1044	1044	1044	1044	1044
Clusters	224	224	224	224	224	224	224	224

	Number Identif.	Number Discrim.	Missing Number	Add	Subtract	Problem Solving	Shape Recogn.	
Panel B: Share of Correct Math Answers by Sub-Domain								
API Original	0.094	0.036	0.099	0.011	0.061	-0.051	0.022	
	[0.252]	[0.661]	[0.192]	[0.874]	[0.402]	[0.481]	[0.789]	
	{0.301}	{0.681}	{0.226}	{0.882}	{0.447}	{0.511}	{0.800}	
	(0.576)	(0.919)	(0.483)	(0.923)	(0.789)	(0.817)	(0.923)	
API Plus	0.259	0.201	0.204	0.215	0.111	0.116	0.099	
	[0.005]	[0.026]	[0.022]	[0.003]	[0.103]	[0.156]	[0.316]	
	{0.011}	{0.036}	{0.035}	{0.008}	{0.130}	{0.200}	{0.365}	
	(0.007)	(0.033)	(0.033)	(0.007)	(0.137)	(0.163)	(0.247)	
API Original = API Plus	[0.095]	[0.103]	[0.218]	[0.008]	[0.500]	[0.046]	[0.396]	
	{0.163}	{0.129}	{0.420}	{0.020}	{0.514}	{0.080}	{0.550}	
	(0.191)	(0.191)	(0.361)	(0.008)	(0.516)	(0.090)	(0.516)	
Observations	1044	1044	1044	1044	1044	1044	1044	
Clusters	224	224	224	224	224	224	224	

Notes: This table shows OLS estimates and the associated *p*-values of the two API modalities: API *Original* and API *Plus* for 1,044 students enrolled in third to sixth grade by the end of the second school year since treatment assignment. For detailed descriptions of the sub-components of the reading and math scores used in this table, see Appendix A.2. The outcome variables are standardized with respect to their means and the standard deviations in the control group. The inference procedures take into account clustering of the error terms at the school level and the block randomization design at the strata level. *p*-values reported in brackets refer to the conventional asymptotic inference. *p*-values reported in braces are computed using randomization inference (randomization-*t*). All *p*-values account for clustering at the school level. *p*-values reported in parentheses are adjusted for testing each null hypothesis (null impact of API *Original*, API *Plus*, and the comparison) on multiple outcomes through the step-wise procedure described in Romano and Wolf (2005a,b, 2016).

Table B-9: Average Program Impacts by the Individual Components of the Socio-Emotional Score

Panel A: First 16 Components

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
API Original	0.040	-0.068	0.074	0.003	-0.008	0.026	0.072	-0.009	0.006	0.015	0.017	0.042	-0.013	-0.024	0.030	-0.020
	[0.293]	[0.041]	[0.049]	[0.943]	[0.835]	[0.477]	[0.047]	[0.818]	[0.863]	[0.679]	[0.646]	[0.205]	[0.737]	[0.410]	[0.348]	[0.563]
	{0.340}	{0.052}	{0.065}	{0.945}	{0.849}	{0.507}	{0.062}	{0.826}	{0.868}	{0.700}	{0.654}	{0.246}	{0.748}	{0.447}	{0.386}	{0.588}
	(0.989)	(0.370)	(0.409)	(1.000)	(1.000)	(0.999)	(0.393)	(1.000)	(1.000)	(1.000)	(1.000)	(0.934)	(1.000)	(0.997)	(0.994)	(0.999)
API Plus	0.125	0.058	0.057	-0.012	-0.014	0.038	0.096	-0.023	0.021	-0.007	0.055	0.056	0.047	0.061	0.040	0.003
	[0.001]	[0.136]	[0.158]	[0.773]	[0.720]	[0.317]	[0.019]	[0.584]	[0.510]	[0.870]	[0.150]	[0.113]	[0.205]	[0.057]	[0.216]	[0.937]
	{0.002}	{0.168}	{0.204}	{0.798}	{0.748}	{0.352}	{0.035}	{0.607}	{0.533}	{0.889}	{0.173}	{0.149}	{0.249}	{0.078}	{0.251}	{0.939}
	(0.010)	(0.775)	(0.813)	(0.999)	(0.999)	(0.972)	(0.157)	(0.997)	(0.995)	(0.999)	(0.809)	(0.710)	(0.901)	(0.421)	(0.908)	(0.999)
API Original = API Plus	[0.044]	[0.002]	[0.690]	[0.721]	[0.863]	[0.777]	[0.560]	[0.739]	[0.696]	[0.595]	[0.380]	[0.706]	[0.141]	[0.014]	[0.759]	[0.532]
	{0.073}	{0.003}	{0.641}	{0.758}	{0.894}	{0.812}	{0.772}	{0.795}	{0.680}	{0.637}	{0.413}	{0.796}	{0.174}	{0.024}	{0.789}	{0.580}
	(0.367)	(0.013)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(0.998)	(1.000)	(0.843)	(0.119)	(1.000)	(0.999)
Observations	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045
Clusters	224	224	224	224	224	224	224	224	224	224	224	224	224	224	224	224

Panel B: Second 16 Components

	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)
API Original	-0.005	-0.050	0.015	-0.030	0.044	-0.034	0.085	-0.026	0.040	0.026	0.060	0.010	0.075	0.002	0.024	0.033
	[0.882]	[0.138]	[0.677]	[0.405]	[0.178]	[0.116]	[0.020]	[0.450]	[0.328]	[0.519]	[0.054]	[0.720]	[0.044]	[0.956]	[0.553]	[0.301]
	{0.894}	{0.159}	{0.707}	{0.448}	{0.192}	{0.143}	{0.038}	{0.491}	{0.370}	{0.564}	{0.076}	{0.730}	{0.067}	{0.967}	{0.564}	{0.345}
	(1.000)	(0.823)	(1.000)	(0.997)	(0.905)	(0.757)	(0.189)	(0.998)	(0.991)	(0.999)	(0.436)	(1.000)	(0.381)	(1.000)	(0.999)	(0.989)
API Plus	0.073	-0.009	0.091	0.021	0.040	-0.013	0.077	0.071	0.045	0.037	0.100	0.053	0.020	0.036	0.037	0.007
	[0.018]	[0.807]	[0.014]	[0.559]	[0.214]	[0.547]	[0.031]	[0.048]	[0.305]	[0.336]	[0.005]	[0.049]	[0.613]	[0.344]	[0.327]	[0.838]
	{0.028}	{0.817}	{0.028}	{0.586}	{0.245}	{0.608}	{0.045}	{0.065}	{0.353}	{0.379}	{0.009}	{0.071}	{0.647}	{0.366}	{0.383}	{0.846}
	(0.154)	(0.999)	(0.117)	(0.997)	(0.908)	(0.997)	(0.258)	(0.371)	(0.972)	(0.972)	(0.037)	(0.379)	(0.997)	(0.972)	(0.972)	(0.999)
API Original = API Plus	[0.018]	[0.246]	[0.055]	[0.191]	[0.923]	[0.350]	[0.848]	[0.012]	[0.925]	[0.796]	[0.301]	[0.193]	[0.203]	[0.422]	[0.735]	[0.494]
	{0.037}	{0.298}	{0.092}	{0.233}	{0.933}	{0.408}	{0.896}	{0.027}	{0.960}	{0.775}	{0.444}	{0.175}	{0.210}	{0.463}	{0.742}	{0.493}
	(0.146)	(0.966)	(0.432)	(0.935)	(1.000)	(0.996)	(1.000)	(0.102)	(1.000)	(1.000)	(0.989)	(0.935)	(0.937)	(0.998)	(1.000)	(0.999)
Observations	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1045	1044
Clusters	224	224	224	224	224	224	224	224	224	224	224	224	224	224	224	224

Notes: This table shows OLS estimates and the associated *p*-values of the two API modalities: API *Original* and API *Plus* for 1,044 students enrolled in third to sixth grade by the end of the second school year since treatment assignment. The individual components of the socio-emotional score are indicator variables for whether the child displays one of the following emotions/behaviors: 1) has serendipitous mood changes, 2) feels or complains that nobody loves him/her, 3) is tense or nervous, 4) lies or cheats, 5) is scared or anxious, 6) talks and argues too much, 7) has difficulty in focusing on a specific activity for an extended amount of time, 8) gets easily confused, 9) it seems that his/her head is in the clouds, 10) threatens or is mean with other children, 11) tends to challenge parental authority, 12) does not feel guilty after a bad deed, 13) does not get along with other children, 14) is impulsive or acts "fast" without thinking, 15) feels has inferiority issues, 16) has no friends, 17) has difficulty letting go certain thoughts, 18) is hyper-active, 19) has a bad temper, or is irascible, 20) looses easily his/her temper, 21) feels unhappy, sad, or depressed, 22) is shy, does not socialize with others, 23) breaks objects on purpose, 24) is too attached to the adults, 25) cries too much, 26) demands a lot of attention, 27) is too much dependent on others, 28) is afraid of other people's judgement, 29) Tends to be in bad company; 30) is reserved, keeps things for himself/herself, 31) worries about every thing, 32) misbehaves at school and does not respect the instructor (see Appendix A.2). The inference procedures take into account clustering of the error terms at the school level and the block randomization design at the strata level. *p*-values reported in brackets refer to the conventional asymptotic inference. *p*-values reported in braces are computed using randomization inference (randomization-*t*). All *p*-values account for clustering at the school level. *p*-values reported in parentheses are adjusted for testing each null hypothesis (null impact of API *Original*, API *Plus*, and the comparison) on multiple outcomes through the stepwise procedure described in Romano and Wolf (2005a,b, 2016).

Table B-10: Comparison of Mentors' Characteristics Across Situations

Variable	(1) Chiapas sample	(2) Experiment 2	(3) Chiapas vs Experiment 2
Male	0.571	0.604	0.033
	(0.495)	(0.491)	[0.492]
Age	28.460	28.266	-0.194
	(3.780)	(3.287)	[0.558]
Speaks Indigenous Language	0.295	0.374	0.079
	(0.457)	(0.486)	[0.101]
Observations	441	139	580

Notes: This table shows the comparison of mentors' characteristics between the second experiment and the scale-up of the *Plus* program. The first two columns show mean and standard deviations (in parentheses) for both samples. The third column shows the difference and the associated *p*-values (in brackets) of the null hypothesis of no difference across samples.

Table B-11: Change in Situation and Impacts on Quality and Quantity of Mentoring Program

	Quantity			Quality		
	Days in Community	Number Activities with Instructor	Time Spent with Instructor	Meetings with Parents of Students at Risk	Students with API Support	Time spent with Students
Change in Situation	-1.585	-1.093	-0.954	-0.596	-0.546	-0.025
	[0.330]	[0.037]	[0.189]	[0.407]	[0.483]	[0.908]
Observations	114	113	114	109	96	110
Observations Survey 2016	56	55	56	51	54	52
Observations Survey 2018	58	58	58	58	42	58

Notes: This table shows the comparison in the quantity and quality of API *Plus* program between the second experiment and the government implementation. This information is collected during the surveys of the local instructors, in the school years 2015-2016 and 2018-2019. Each estimate in each column represents an OLS estimate for the difference in the mentoring services across the two situations. The asymptotic *p*-values are reported in square brackets. All the regressions include the same set of controls as in Table 4.

Table B-12: The Impact of the API Plus Program on School Closures

	Non-Experimental Schools	Experimental Schools
API Plus	-0.068	-0.070
	[0.000]	[0.026]
Observations	1161	184

Notes: This table shows the OLS estimates of the assignment to the API program during the government implementation of the Plus modality on the rate of school closures as measured over the subsequent two years. *p*-values reported in brackets are based on asymptotic inference. All the regressions include the same set of controls as in Table 4.

Table B-13: Average Program Impacts by the Individual Components of Parental Investments

	Engage with School				Manage School Resources					Engage with Child			
	Volunteering	Donate Cash	Donate In-Kind	Food Instructor	Manage School Resources	Propose School Material	Decide School Material	Decide Money Allocation	Evaluate School Targets	Help With Homework	Extra-Academic Activities	Meeting Teachers	Expect Upper Secondary
							Panel A: First Experiment						
API Original	0.042	0.118	0.063	0.046	-0.042	0.026	-0.009	0.002	-0.040	0.210	0.055	0.203	0.025
	[0.417]	[0.126]	[0.478]	[0.560]	[0.579]	[0.726]	[0.912]	[0.974]	[0.487]	[0.358]	[0.528]	[0.291]	[0.608]
	{0.435}	{0.147}	{0.494}	{0.566}	{0.597}	{0.734}	{0.916}	{0.971}	{0.512}	{0.382}	{0.524}	{0.322}	{0.626}
	(0.955)	(0.475)	(0.969)	(0.969)	(0.969)	(0.969)	(0.983)	(0.983)	(0.969)	(0.928)	(0.969)	(0.872)	(0.969)
Number of clusters	73	73	73	73	73	73	73	73	73	73	73	73	73
Observations	208	208	207	208	208	208	208	208	208	208	207	208	199
							Panel B: Second Experiment						
API Original	-0.031	-0.004	-0.058	-0.058	-0.029	-0.070	-0.062	-0.010	-0.027	0.222	0.074	0.043	0.010
	[0.356]	[0.894]	[0.130]	[0.042]	[0.471]	[0.095]	[0.122]	[0.772]	[0.389]	[0.027]	[0.082]	[0.568]	[0.781]
	{0.387}	{0.898}	{0.159}	{0.067}	{0.516}	{0.124}	{0.163}	{0.793}	{0.438}	{0.049}	{0.126}	{0.589}	{0.806}
	(0.884)	(0.981)	(0.452)	(0.194)	(0.917)	(0.369)	(0.452)	(0.981)	(0.888)	(0.137)	(0.350)	(0.942)	(0.981)
API Plus	0.036	0.018	0.044	0.071	0.069	0.001	0.006	0.010	0.018	0.221	0.108	0.192	0.094
	[0.289]	[0.625]	[0.329]	[0.013]	[0.095]	[0.978]	[0.890]	[0.776]	[0.570]	[0.066]	[0.015]	[0.020]	[0.019]
	{0.323}	{0.663}	{0.372}	{0.020}	{0.130}	{0.981}	{0.891}	{0.784}	{0.597}	{0.092}	{0.023}	{0.031}	{0.029}
	(0.765)	(0.953)	(0.778)	(0.062)	(0.323)	(0.977)	(0.977)	(0.977)	(0.953)	(0.245)	(0.063)	(0.072)	(0.072)
Original = Plus	[0.092]	[0.578]	[0.026]	[0.000]	[0.035]	[0.142]	[0.139]	[0.631]	[0.210]	[0.999]	[0.475]	[0.107]	[0.047]
	{0.096}	{0.610}	{0.045}	{0.000}	{0.050}	{0.159}	{0.163}	{0.634}	{0.231}	{0.756}	{0.563}	{0.147}	{0.084}
	(0.363)	(0.877)	(0.124)	(0.005)	(0.158)	(0.441)	(0.441)	(0.877)	(0.534)	(0.998)	(0.875)	(0.389)	(0.199)
Observations	1042	1042	1039	1042	1033	1036	1027	1031	1029	1044	1033	974	1017
Clusters	224	224	224	224	224	224	224	223	224	224	224	223	224

Notes: This table shows OLS estimates and the associated *p*-values of the two API modalities: API *Original* and API *Plus* for 1,044 students enrolled in third to sixth grade by the end of the second school year since treatment assignment. For a detailed descriptions of the sub-components of the reading and math scores used in this table, see Appendix A.2. The outcome variables are standardized with respect to their means and the standard deviations in the control group. The inference procedures take into account clustering of the error terms at the school level and the block randomization design at the strata level. *p*-values reported in brackets refer to the conventional asymptotic inference. *p*-values reported in braces are computed using randomization inference (randomization-*t*). All *p*-values account for clustering at the school level. *p*-values reported in parentheses are adjusted for testing each null hypothesis (null impact of API *Original*, API *Plus*, and the comparison) on multiple outcomes through the stepwise procedure described in Romano and Wolf (2005a,b, 2016).

Table B-14: Remedial Education Sessions

	Reading Score	Math Score	Socio-Emotional Score	Overall Index
API *Original*× Rank≥7	0.193	0.023	0.147	0.192
	[0.105]	[0.844]	[0.313]	[0.177]
API *Plus*× Rank≥7	0.423	0.274	0.206	0.430
	[0.001]	[0.055]	[0.140]	[0.003]
API *Original*× Rank<7	0.078	0.045	0.034	0.074
	[0.431]	[0.641]	[0.728]	[0.487]
API *Plus*× Rank<7	0.261	0.224	0.183	0.327
	[0.011]	[0.042]	[0.082]	[0.003]
H0: *Original* = *Plus* (<7)	[0.104]	[0.095]	[0.192]	[0.039]
H0: *Original* = *Plus* (≥7)	[0.072]	[0.081]	[0.721]	[0.144]
H0: [*Original-Plus* (<7)]=[*Original-Plus* (≥7)]	[0.766]	[0.675]	[0.639]	[0.937]
Number of Schools	224	224	224	224
Number of Observations	1044	1044	1045	1045

Notes: This table shows the estimates for the API program once we interact the treatment assignment dummies with indicators of whether a child is among the six lowest-performing children in the class on the diagnostic test (Rank Below 7 and Rank Above 7), which is one of the main determinants for participation in the one-on-one remedial sessions with the mentors (see Figure B-3). Reading, math, and socio-emotional scores are standardized with respect to the mean and the standard deviation of the control group. See Appendix A.2 for a detailed description of the outcome variables. Asymptotic p-values reported in brackets are clustered at the school level.

Table B-15: Teacher Pedagogical Practices

	Learning Activities	Engage With Students	Manage Time	Use of Material	Overall Index
API *Original*	0.047	-0.067	0.095	-0.127	-0.038
	[0.722]	[0.671]	[0.635]	[0.465]	[0.789]
	{0.723}	{0.666}	{0.636}	{0.458}	{0.799}
	(0.966)	(0.966)	(0.966)	(0.923)	(0.966)
API *Plus*	-0.072	0.051	-0.084	0.023	-0.202
	[0.616]	[0.733]	[0.627]	[0.877]	[0.150]
	{0.605}	{0.734}	{0.641}	{0.880}	{0.145}
	(0.949)	(0.949)	(0.949)	(0.949)	(0.278)
Original = Plus	[0.464]	[0.481]	[0.357]	[0.425]	[0.277]
	{0.448}	{0.501}	{0.385}	{0.414}	{0.277}
	(0.707)	(0.707)	(0.707)	(0.707)	(0.591)
Observations	209	209	209	209	209

Notes: This table shows OLS estimates and the associated p-values of the API *Original* and the API *Plus* modalities on teachers' pedagogical practices (Stallings Classroom Snapshot). The outcome variables are standardized with respect to their means and the standard deviations in the control group. The inference procedures take into account the block randomization design at the strata level. p-values reported in brackets refer to the conventional (robust) asymptotic inference. p-values reported in braces are computed using randomization inference (randomization-t). p-values reported in parentheses are adjusted for testing each null hypothesis (null impact of API *Original*, API *Plus*, and the comparison) on multiple outcomes through the stepwise procedure described in Romano and Wolf (2005a,b, 2016).

References

Bruns, B. and J. Luque (2015). *Great Teachers : How to Raise Student Learning in Latin America and the Caribbean.* Washington, DC: World Bank.

Dubeck, M. M. and A. Gove (2015). The early grade reading assessment (egra): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development 40*, 315–322.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics 40*(4), 1079–1087.

Platas, L. M., L. R. Ketterlin-Geller, and Y. Sitabkhan (2016). Using an assessment of early mathematical knowledge and skills to inform policy and practice: Examples from the early grade mathematics assessment. *International Journal of Education in Mathematics, Science and Technology 4(3)*, 163–173.

Romano, J. P. and M. Wolf (2005a, March). Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing. *Journal of the American Statistical Association 100*, 94–108.

Romano, J. P. and M. Wolf (2005b, July). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica 73*(4), 1237–1282.

Romano, J. P. and M. Wolf (2016). Efficient Computation of Adjusted p-Values for Resampling-Based Stepdown Multiple Testing. *Statistics & Probability Letters 113*(C), 38–40.