I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14192

Enhancing Human Capital at Scale

Francesco Agostinelli
Ciro Avitabile
Matteo Bobba

MARCH 2021

DISCUSSION PAPER SERIES

IZA DP No. 14192

Enhancing Human Capital at Scale

Francesco Agostinelli
University of Pennsylvania

Ciro Avitabile
World Bank

Matteo Bobba
University of Toulouse Capitole and IZA

MARCH 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Enhancing Human Capital at Scale*

A two-year randomized evaluation shows that the effectiveness of multi-tasking mentors on schooling outcomes crucially depends on their training. While a standard training modality in highly marginalized communities in Mexico generates null results, enhanced training yields sizable treatment effects on primary school children's cognitive, behavioral, and educational achievements. This difference cannot be explained by remedial educational activities or pedagogical support, but it can be reconciled with higher parental aspirations and investments. Evidence gathered on the subsequent national roll out of the intervention with enhanced training substantiates the external validity of our findings.

JEL Classification: H43, I10, I20, I38

Keywords: learning outcomes, cognitive achievement, pedagogical practices, parental inputs, scaling-up effects

Corresponding author:
Matteo Bobba
Toulouse School of Economics
University of Toulouse Capitole
21 Allée de Brienne
31015 Toulouse
France

E-mail: matteo.bobba@tse-fr.eu

* We thank the Consejo Nacional de Fomento Educativo (CONAFE) for the generous collaboration throughout this project and Alonso Sanchez for his initial input into the project. Miguel Angel Monroy provided excellent research assistance. This study is registered in the AEA RCT Registry and the unique identifying number is AEARCTR-0001645. Avitabile acknowledges financial support for data collection from the Strategic Impact Evaluation Fund (SIEF) of the World Bank and the Consejo Nacional de Evalución de la Politica de Desarollo Social (CONEVAL). Bobba acknowledges financial support from the AFD, the H2020-MSCA-RISE project GEMCLIME-2020 GA No 681228, and the ANR under grant ANR-17-EURE- 0010 (Investissements d'Avenir program).

1 Introduction

There is a wide learning gap among school-aged children between urban and rural areas in developing countries (World Bank, 2018). Efforts to improve education outcomes in rural areas are often hampered by low levels of student readiness (Gertler et al., 2014), high rates of teacher absenteeism and turnover (Duflo et al., 2012; Albornoz et al., 2020), and inadequate parental support (Attanasio et al., 2020a).

In this paper, we provide new insights on whether and how a multifaceted approach can improve child outcomes in remote areas of Chiapas—the poorest state of Mexico. In the program, the government enlists recent university graduates as educational mentors in an attempt to improve the quality of schooling in disadvantaged communities. These young professionals provide pedagogical support to the local instructors, organize one-on-one remedial education sessions for the students who are lagging behind, and encourage parental involvement in children's education through home visits. The evidence is drawn from a field experiment recently implemented in close collaboration with the government agency that is in charge of delivering education services in such low-resource environments. We randomly vary both the presence and the training intensity of the mentors among a subset of multi-grade primary schools.

Our main result is that training matters for the effectiveness of the educational mentors in our context. While a low-intensity training modality yields null results on child outcomes after two years of exposure when compared to the control group with no mentors, enhanced training—in the form of hands-on strategies to teach basic reading and math competencies as well as periodic peer-to-peer meetings during the school-year—leads to sizable gains in children's reading scores (+0.30 standard deviations), math scores (+0.24 standard deviations), socio-emotional scores (+0.20 standard deviations), and a marginally significant effect on the probability of enrolling in lower secondary education (+8.8 percentage points, out of a basis of 63 percent enrollment in the control group).

The large difference in effect sizes between the two training modalities does not seem to be explained by either changes in the pedagogical practices of the community instructors or by differences in the relative effectiveness of the remedial educational sessions. Instead, we provide evidence that mentors with enhanced training engage more with parents, in terms of both the quantity and the quality of their periodic interactions. This mechanism is corroborated by evidence of differential impacts of the training modality on parental aspirations toward children's education attainments as well as different measures of parental investments.

In recent years, scholars and policy makers alike have been increasingly concerned about the ability of randomized controlled trials (RCTs) to inform policy decisions, since interventions that have been found effective in small-scale pilots often fail to hold to their promises when implemented at scale (Bold et al., 2018; Cameron et al., 2019; Muralidharan and Singh, 2020). The subsequent national scale-up of the enhanced training modality gives us the unique opportunity to study the effectiveness of the program *for the same set of communities* under both the small-scale experimental regime and the large-scale policy environment. We find that one year of exposure to the intervention at scale increases the probability of children's transitioning into secondary school by 9.4 percentage points, which is remarkably close to the experimental estimate.[1] We also find that schools are more likely to remain open and to continue their service in the community two years after the end of the experiment. This result is consistent with the hypothesis that educational investments at the community level are a socially determined outcome (List et al., 2019).

Our findings contribute to previous literature by showing that in a low-resource environment, a scalable intervention aimed at improving the quality of education provision can be effective in the production of children's human capital (Fryer et al., 2015; Carneiro et al., 2019, 2020; Andrew et al., 2020).[2] More closely related to our paper, Attanasio et al. (2014, 2020b) document that home visits programs can spur cognitive and socio-emotional outcomes for children in the first years of life through increased parental involvement. Our results are consistent with the notion that teachers' pedagogical practices are difficult to improve, especially in disadvantaged contexts (Yoshikawa et al., 2015; Ozler et al., 2016; Bassi et al., 2020). While there is evidence on the effectiveness of remedial education interventions for under-performing children (Banerjee et al., 2007, 2017), we show that this is not explained by the role of the tutor's training.

2 Context and Design of the Evaluation

2.1 *CONAFE and the Mobile Pedagogical Mentors*

CONAFE is a semi-autonomous government agency responsible for providing schooling services in highly marginalized areas of Mexico with a significant share of indigenous population. In those communities, CONAFE offers all education services from pre-school until the end of lower secondary school (9th grade). Primary schools typically have a single multi-grade classroom with 10–15 students, where a community resident with a secondary school degree acts as an instructor without any formal qualification.

In 2009, CONAFE launched the "Mobile Mentors" (*Asesores Pedagogicos Itinerantes,* API henceforth) program as an attempt to improve the quality of education provision in schools located in remote and disadvantaged areas. The mentors are selected from recent university graduates. Preference is given to applicants with degrees in pedagogy, psychology, sociology and social services who have previous experience as community instructors and who speak

[1]The program under study has indeed many features of a "scalable" intervention (Banerjee et al., 2017; Muralidharan and Niehaus, 2017; Al-Ubaydli et al., 2019). First, the field experiment has been implemented by the same agency that was in charge of the program scale-up. Second, the implementing agency and the research team designed the training modalities of the experiment bearing in mind the financial and human resources constraints at scale. Third, the overall scale of the program operation remained the same before, during, and after the experiment.

[2]There is also a consensus that gaps in family investments are behind the gaps in children's achievements among different socioeconomic groups (Heckman and Mosso, 2014).

an indigenous language.[3] They are usually hired for a two–year period and receive a monthly salary of MXN $6,000 (USD $345 in 2015).[4] They alternate on two-week intervals between two nearby school communities during two consecutive academic years. More information about the API intervention is provided in Appendix A.2.

The mentors carry out three main activities in each school community with a predetermined time allocation: (i) one-on-one tutoring to the least-performing students in remedial sessions (60 percent of their time), (ii) pedagogical support to teachers (15 percent), and (iii) visiting parents at their homes to provide them with information on their children's progress in school and promote their participation in school activities (25 percent). Each mentor is assigned to a maximum of six students for the personalized remedial sessions, which in principle should take place outside of regular school hours. During the regular school hours, the mentor is supposed to observe and take notes about the teaching practices of the local instructor, help them with the students who have learning difficulties, and work outside the classroom with those students who cannot attend one-on-one tutoring in the afternoon. In addition to working on behavioral issues directly with the children, the mentors are supposed to address them with parents as part of the home visits.

2.2 *Evaluation Design and Data*

In close collaboration with CONAFE, we developed and unrolled a field experiment with the objective of evaluating the relative effectiveness of one modality aimed at strengthening the API program. The API Standard modality is meant to track the benchmark intervention described in Section 2.1. The API Plus modality embeds all the features of the API Standard, with a significant change in the training module. The module entails two weeks rather than one week of classes, with the second week focused on hands-on strategies to teach basic reading and math competencies. In addition, mobile mentors attend four peer-to-peer training sessions during the school year. These three-day sessions (18 hours per session) allow participants to share experiences and design common strategies to better organize their day-to-day activities in the communities.

We randomly select 230 schools in rural Chiapas from a set of eligible schools (see Appendix A.2) that were not previously part of the API program.[5] The assignment of the program treatment is carried out using a randomized block design clustered at the school level, with the strata represented by the deciles of the 2012 school average in a national standardized achievement score. As a result, 60 schools are assigned to the API Plus, 70 schools are assigned to the API Standard, and the remaining 100 schools are in the control group with no API intervention.

[3] CONAFE advertises the program both with on-campus visits and announcements through the media.

[4] Partly due to the fixed-term contract, they are not covered by either social security or healthcare assistance. Teachers under the regular system have a permanent contract with a monthly salary starting at MXN$ 8,000 and have both healthcare and social security benefits.

[5] The communities of the schools in this sample are very small, averaging 100 inhabitants, and they are difficult to access, with one-fifth of the schools having no road access whatsoever.

The intervention was rolled out in August 2014. The first round of data collection took place in the spring of 2016. By that time, six schools closed either temporarily or permanently, one quarter of the community instructors reported eight or fewer months of tenure in the school, and only 48 out of the original 130 mobile mentors were working in the same schools where they had been originally assigned. All these outcomes are well-balanced across treatment arms.[6] The CONAFE delegation in Chiapas arranged for a replacement within two weeks from the day of the API departure from a community. A follow-up survey was conducted in the fall of 2018. Additional design and implementation details of the evaluation are discussed in Appendix B while a full description of the different datasets employed in the analysis is provided in Appendix C.

The main sample of the analysis is comprised of 224 schools and 1,045 students enrolled in 3rd to 6th grade with complete records of our outcomes of interest.[7] Appendix Table E.1 shows the balance with respect to treatment assignment for a large set of predetermined covariates. Characteristics are well balanced across the three groups, except for a larger share of indigenous in the Plus group, which is statistically significant at the 10 percent level.

Households in our sample are on average very poor, with 80 percent of them report being *Oportunidades* beneficiaries and only one fourth having access to the sewage system. About two thirds of the primary care givers of the children in our sample did not complete primary education, and almost 30 percent are illiterate. Consistently with the targeting design of the API program, the majority (80 percent) of the mentors in our sample have previous experience in CONAFE schools as community instructors. For the schools in our sample, the average standardized scores in math and Spanish are 0.5 and 0.7 standard deviations below the national averages.[8] During the school years 2014–15 and 2015–16, less than two thirds of the sixth graders in the control group enroll in secondary schooling while the national average is 95.2 percent.

[6]Two schools were found closed in the control group, two in the API Standard group, and two in the API Plus group. The p-values of the Komolgorov-Smirnoff statistic for the equality of the distributions of school tenure of the community instructors in each treatment arm and the control group are 0.773 and 0.892, respectively. The p-value of the Plus-Standard difference in the share of mentors who drop-out from the API program during the experiment is 0.957. There is no evidence of composition changes between the Standard and Plus groups induced by the API turnover (see Table E.3 in the Appendix).

[7]We administered the standardized achievement test only for students enrolled in 3rd to 6th grade. See Appendix C.2 for further details on the measures of cognitive achievement used throughout the analysis.

[8]The achievement tests were administered to all primary school students in Mexico until 2013.

3 Experimental Evidence

3.1 *Standard versus Plus: Main Impacts*

The causal effects of the two API treatments can be estimated via the following regression model

$$Y_{i,j} = \beta_0 + \beta_1 Standard_j + \beta_2 Plus_j + \gamma' X_{i,j} + u_{i,j} \,, \tag{1}$$

where $Y_{i,j}$ is an outcome of student i in school j, which is recorded by the end of the second consecutive school year since the assignment of the mobile mentors to the communities. The two variables $Standard_j$ and $Plus_j$ take a value of one if school j is assigned to either the API Standard or the API Plus group, respectively. Both variables take a value of zero if school j is assigned to the control group. The vector $X_{i,j}$ consist of individual and community–level characteristics. Individual-level characteristics include gender and an indicator for whether the child speaks an indigenous language (see Appendix Table E.1). Community-level characteristics include the strata indicators that account for the block randomization design, a set of indicators for survey weeks and survey routes,[9] as well as the school average scores in Math, Spanish, and Science as measured in 2013. Standard errors are clustered at the school level so as to account for correlated shocks that vary at the same level as the treatment indicators.

We have different outcomes of interest. Cognitive achievement of children is measured via the math and reading scores. Socio-emotional skills are measured via a behavioral problem index.[10] To better interpret our results, we standardize all the test scores using the mean and the standard deviation observed in the control group. Finally, we consider an indicator variable for whether the child has transitioned to a secondary school.

Table 1 displays the main estimates of the impact of the API program. When compared to the control group, children who are enrolled in a school that received the Standard API model increased their reading scores by 0.13 standard deviations, as opposed to a 0.32 standard deviations improvement for those attending schools served by an API Plus. While the estimated effect of the API Standard on reading scores is not statistically different from zero, we can reject the hypothesis of a null effect of API Plus at the 99 percent confidence level. Quantitatively, the API Plus effect is approximately 2.5 times higher than the API Standard effect. This difference is statistically different from zero (p-value =0.043).[11]

[9] During the data collection, a few schools had to be surveyed on a second or third visit due to adverse weather conditions or high political instability.

[10] See Appendix C.2 for a detailed description of the cognitive and socio-emotional scores used in the analysis. We measure a child's reading and math skills via the Early Grade Reading Assessment (EGRA) and the Early Grade Math Assessment (EGMA), respectively. The socio-emotional score is re-scaled in such a way that higher values are associated with fewer behavioral issues.

[11] In Table E.5 we test three possible channels behind possible changes in classroom composition: grade retention, inflows of new students from CONAFE, and inflows of students from outside the CONAFE system. We do not find any difference among the treatment arms in any of the these outcomes.

Table 1: Average Program Impacts

	Reading Score	Math Score	Socio-emotional Score	Transition to Secondary School
API Standard	0.126	0.056	0.071	0.075
	(0.077)	(0.075)	(0.087)	(0.049)
API Plus	0.315	0.237	0.199	0.089
	(0.083)	(0.089)	(0.087)	(0.047)
H0: Standard=Plus	0.043	0.043	0.178	0.777
Mean Control Group	0.000	0.000	0.000	0.617
SD Control Group	1.000	1.000	1.000	0.487
Observations	1044	1044	1045	992
Clusters	224	224	224	223

Notes: This table shows the estimates of the two API modalities: API Standard and API Plus. Each coefficient represents the estimated effect of the program relative to the control group as depicted in the regression model (1). Reading, Math, and Socio-emotional scores are standardized with respect to the mean and the standard deviation of the control group. The last column is a linear probability model for the sub-set of students enrolled in the sixth grade, where the transition to Secondary School is an indicator variable of whether or not students enroll in a secondary school during the next school year. See Appendix C for a detailed description of the outcome variables. Standard errors reported in parentheses are clustered at the school level.

We find similar patterns when we look at math scores. On the one hand, the API Standard had a small (0.06 standard deviations) and not statistically significant effect. On the other hand, we find a sizable effect of API Plus, with an estimated treatment effect of 0.24 standard deviations. This effect is statistically significant, and we can reject the hypothesis that the two treatment arms have the same effect at the 95 percent confidence level.[12]

We also find that the assignment of the API Standard produces a small improvement of 0.07 standard deviations in a child's socio-emotional skills, although the estimated effect is not statistically different from zero. On the contrary, the API Plus group generates a sizable improvement in the socio-emotional score of 0.2 standard deviations. This larger effect for the API Plus is consistent with qualitative evidence documenting that mentors in the API Plus group shared more effective strategies to best deal with children's emotions during the bimonthly peer-to-peer sessions.

The last column in Table 1 reports the estimated effects on the transition to secondary school. In this case, the outcome is a dummy variable for whether or not children enrolled in secondary school for the subset of children who were enrolled in 6th grade during the school years 2014–15 and 2015–16. The API Plus treatment increases the probability of a child's enrolling in a lower secondary school by 8.9 percentage points. Although marginally

[12]In Table E.2 we report the results by sub-domains of the reading scores (panel A) and math scores (panel B).

Table 2: Quantile Program Impacts

Panel A: Reading Score

	10th Perc.	30th Perc.	50th Perc.	70th Perc.	90th Perc.
API Standard	0.181	0.135	0.159	0.061	0.008
	(0.179)	(0.133)	(0.093)	(0.056)	(0.039)
API Plus	0.288	0.306	0.295	0.094	0.065
	(0.212)	(0.154)	(0.098)	(0.057)	(0.048)
H0: Standard=Plus	0.673	0.239	0.199	0.627	0.283
Mean Control at Percent.	-1.673	-0.496	0.377	0.704	1.031
Observations	1044	1044	1044	1044	1044
Clusters	224	224	224	224	224

Panel B: Math Score

	10th Perc.	30th Perc.	50th Perc.	70th Perc.	90th Perc.
API Standard	0.137	0.044	0.073	0.007	0.044
	(0.175)	(0.140)	(0.115)	(0.085)	(0.089)
API Plus	0.312	0.270	0.213	0.152	0.155
	(0.143)	(0.149)	(0.125)	(0.104)	(0.097)
H0: Standard=Plus	0.313	0.151	0.306	0.180	0.231
Mean Control at Percent.	-1.346	-0.555	0.111	0.690	1.230
Observations	1044	1044	1044	1044	1044
Clusters	224	224	224	224	224

Panel C: Socio-Emotional Score

	10th Perc.	30th Perc.	50th Perc.	70th Perc.	90th Perc.
API Standard	-0.058	0.096	0.041	0.058	0.136
	(0.103)	(0.097)	(0.106)	(0.164)	(0.202)
API Plus	0.022	0.120	0.191	0.270	0.147
	(0.114)	(0.118)	(0.098)	(0.168)	(0.186)
H0: Standard=Plus	0.517	0.835	0.212	0.186	0.964
Mean Control at Percent.	-1.329	-0.554	-0.088	0.532	1.462
Observations	1045	1045	1045	1045	1045
Clusters	224	224	224	224	224

Notes: This table shows the estimates of quantile regressions for reading, math, and socio-emotional scores at different quantile levels (columns). The omitted category refers to the control group. See Appendix C for a detailed description of the outcome variables. Standard errors in parentheses are clustered at the school level (Parente and Santos Silva, 2016).

significant, the effect is quantitatively sizable, as it represents a 14 percent increase in the share of students who transit to secondary school for the API Plus group relative to the mean in the control group. As for the rest of the outcomes displayed in Table 1, the effect of the API Standard (an increase of 7.5 percentage points) is not statistically different from zero.

Table 2 presents treatment effect estimates at various percentiles of the test score distributions. Results show a more pronounced impact of both API Standard and API Plus on reading and math achievement for pupils who perform relatively worse on those achievement tests (panels A and B). Effect sizes for the API Plus are larger throughout the achievement distribution, although we cannot reject that they are equal to those of the API Standard.

These heterogeneous effects imply a substantial reduction in the within-class dispersion of cognitive achievements due to the API Plus.[13] Estimates are more erratic for the socio-emotional score (panel C).

Overall, the results presented in Tables 1 and 2 show that the API intervention had differential impacts according to the training received by the mobile mentors. While the API Standard modality did not significantly boost any of the outcomes of interest, the API Plus modality had sizable effects on children's cognitive and socio-emotional scores, as well as on schooling attainments.

3.2 *Standard versus Plus: Potential Mechanisms*

As mentioned in Section 2.1, the activities of the pedagogical mentors are organized around three main areas of intervention: (i) one-on-one tutoring sessions with academically weaker students, (ii) pedagogical support to teachers, and (iii) parental engagement through periodic home visits and school meetings. In this section, we leverage detailed survey information to provide direct evidence on the possible channels behind the differential impacts between the API Plus and the API Standard on a child's development.

The first two columns of panel A in Table 3 present the average impacts of the two treatments on parental outcomes. These variables are constructed from the household survey collected in the spring of 2016 (see Appendix C.2). The estimated coefficients reported in the first column show that the API Plus modality seems effective in boosting parental expectations about children's educational achievement. The point estimate implies that parents are 9.3 percentage points more likely to expect their child to complete secondary schooling, which represents a 12 percent increase with respect to the sample mean in the control group. The corresponding effect size for the API Standard is very small in magnitude and not statistically different from zero—we can reject the null hypothesis of equal treatment effects at the 95 percent confidence level. A similar asymmetric response emerges when we consider a parental investment index (second column), generated as the principal component of five different measures of educational investments, such as helping with homework and participating in school activities. Parents who have been exposed to the API Plus indeed seem to be significantly more engaged in their children's education.[14]

We next focus on the pedagogical practices of the community instructors, which draw from the classroom observation survey (see Appendix C.2). The last two columns in panel A of

[13] Figure E.1 in the Appendix plots the empirical frequencies of the class-level standard deviation for reading, math, and socio-emotional scores (panels A, B and C, respectively) as well as for a class-level entropy index of whether a child has enrolled in secondary schooling (panel D). Appendix Table E.4 shows the OLS estimates of school-level regressions for the effects of the API Standard and the API Plus on the same four outcomes. While estimates are noisy and relatively small in magnitudes for the API Standard, the API Plus leads to a 0.1 decrease in the dispersion of both cognitive measures and transitions.

[14] We also estimate treatment effects of both the API Standard and API Plus treatments for each of the individual measures of the parental investment index. The results are reported in Appendix Table E.7 and they are broadly consistent with the overall index.

Table 3: Standard vs. Plus—Parents and Community Instructors

Panel A: Parental Inputs and Pedagogical Practices				
	Parent Expect Complete Sec.	Parental Investment Index	Instructor is Out of the Class	Instructor Keeps Rhythm
API Standard	0.010	-0.070	-0.996	0.067
	(0.036)	(0.091)	(0.769)	(0.032)
API Plus	0.093	0.239	-1.069	0.086
	(0.039)	(0.099)	(0.563)	(0.043)
H0: Standard=Plus	0.044	0.009	0.896	0.656
Mean Control Group	0.755	-0.002	2.452	0.018
SD Control Group	0.431	1.001	7.344	0.133
Observations	1016	963	259	252
Clusters	224	222	.	.

Panel B: Mentors' Interactions with Parents				
	Meetings in Last 60 Days	Visits in Last 60 Days	Inform About Child	Advise About Child
API Plus	1.015	0.688	0.105	0.100
	(0.711)	(0.468)	(0.051)	(0.045)
Mean API Standard	5.037	3.039	0.714	0.749
Observations	482	491	354	353
Clusters	123	124	113	112

Notes: This table shows the estimates for the behavioral responses in parental investments, instructors' pedagogical practices, and mentor-parent interactions. The first two columns in panel A show the estimates for parental aspirations and parental investments taken from the household survey. The last two columns in panel A show the estimates for two measures of teaching practices taken from an adapted version of the Stallings classroom snapshot (Bruns and Luque, 2015). For each model in panel A, the omitted category refers to the control group. Panel B shows the estimates for the mentor-parent interactions as reported by the parents in the household survey. See Appendix C.2 for a description of the two surveys, which were both administered in the spring of 2016. For each regression in panel B, the omitted category refers to the API Standard group (the control group does not receive home visits according to the experimental design). Standard errors in parentheses are clustered at the school level for parental outcomes and mentor-parent interactions.

Table 3 report estimates of the effect of the two API modalities using data at the school level. The estimated coefficients in the third column show that the presence of the API leads to a reduction in teachers' unjustified absence by roughly one minute in a representative hour of teaching time in both modalities. Albeit small in absolute terms and noisily estimated, the relative effect size is about half of the mean in the control. The estimates shown in the fourth column reveal a positive effect of the two API modalities on an indicator variable for whether the instructor is able to keep the rhythm of the class while teaching. The estimated effects are sizable when compared to the very low share of teachers in the control group who are effective in this pedagogical dimension. For both instructors' outcomes, the effects are quantitatively and statistically similar across API modalities.[15]

The asymmetric effects on parental outcomes are consistent with the estimates reported in panel B of Table 3, which show that the API Plus modality improved the quantity (columns 1 and 2) and the quality (columns 3 and 4) of the mentor-parent interactions relative to the API Standard. On average and over a two-month period, mobile mentors in the Plus group meet one time more with parents at school and 0.7 times more at home (sample means in the API Standard group are 5 and 3, respectively) compared to those in the Standard modality, although the effects are noisily estimated. The last two columns of panel B show more-precise estimates on two outcomes: (i) an indicator variable for whether the mentors have informed the parents about their children's learning difficulties; and (ii) whether the mentors provide concrete advice to parents on how to tackle these difficulties (column 4). For both outcomes, the effect sizes imply a 14 percent increase in the probability of informing parents with respect to the sample means in the API Standard group.

Last, we evaluate the role of the tutoring sessions in potentially explaining the differential treatment effects of the API Standard and the API Plus. As mentioned in Section 2.1 (see Appendix A.2 for more details), those take the form of individual meetings between each of the mentors and the students outside of regular school hours. The six weakest students in the class are deemed eligible for these extra tutoring sessions, based on three criteria: (i) the score in a diagnostic exam in Spanish, math, and natural science, (ii) the assessment of the community instructor, and (iii) an ad hoc diagnostic test applied by the mentor. We only have information on the first eligibility criterion for all the schools in our sample. Hence, we can use the average diagnostic score across the three subjects as a predictor for the child's participation in the tutoring sessions.[16]

We consider a simple variant of the regression model depicted in equation (1) in which the API Standard and API Plus categories are interacted with indicator variables for whether each child is among the six weakest students—i.e. the reverse rank is smaller than 7.[17] The estimates reported in Table 4 suggest that remedial education cannot explain the difference in the impacts between the two API modalities. Effect sizes for the reading and the math scores are reported in the first two columns, and they confirm the larger and significant effect of the API Plus modality when compared to the API Standard. Importantly, the

[15]In Appendix Table E.6 we complement these instructors' outcomes with additional measures of pedagogical practices. The results show erratic patterns with no significant effects of neither API modalities.

[16]We do not know the weight given to each of these three subjects. For this reason we use the average between the three diagnostic scores. We restrict the sample to treatment schools and we run a Probit model of the relationship between the probability of participating in the tutoring sessions and the class-level reverse ranking of students as implied by their average score in the three subjects, with the worst-performing student being ranked first and so on. The estimated marginal effects are plotted in Figure E.2 in the Appendix. We observe a statistically significant drop in the predicted probability of participation in the tutoring sessions for all students who are ranked seventh or above in the average score for the three subjects. We observe a similar drop when conducting the same exercise for each subject separately (results available upon requests).

[17]Individual and household characteristics are balanced for the subsamples of children who are below and above rank 7 (see Tables E.8 and E.9 in the Appendix). In Appendix D, we provide more details on the interpretation of the results reported in Table 4.

Table 4: Standard vs. Plus—Remedial Education Sessions

	Reading Score	Math Score	Socio-Emotional Score
API Standard×Rank Above 7	0.209	0.044	0.163
	(0.122)	(0.117)	(0.147)
API Plus×Rank Above 7	0.436	0.292	0.223
	(0.129)	(0.146)	(0.139)
API Standard×Rank Below 7	0.037	-0.003	0.045
	(0.098)	(0.097)	(0.099)
API Plus×Rank Below 7	0.262	0.228	0.196
	(0.098)	(0.105)	(0.105)
Rank Below 7	-0.088	-0.180	0.100
	(0.100)	(0.102)	(0.113)
H0: Standard=Plus (Below 7)	0.045	0.035	0.185
H0: Standard=Plus (Above 7)	0.079	0.087	0.708
Observations	1044	1044	1045
Clusters	224	224	224

Notes: This table shows the estimates for the linear regression model depicted in equation (1) 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 Appendix Figure E.2). Reading, math, and socio-emotional scores are standardized with respect to the mean and the standard deviation of the control group. See Appendix C.2 for a detailed description of the outcome variables. Standard errors in parentheses are clustered at the school level.

magnitude of the differences between API modalities is very comparable between students who are ranked sixth or below—and hence who are more likely to be targeted by the remedial education sessions (see Appendix Figure E.2)—and for those who are ranked seventh or above. Consistent with the evidence reported in panel C of Table 2, the estimates are more erratic for the socio-emotional score and no clear patterns emerge.[18]

In summary, while remedial education sessions or pedagogical support from the mentors are unlikely to explain the asymmetric impacts of the two API modalities, the increased mentor-parent interactions, which may have triggered higher parental aspirations and investments in children, can potentially explain these results.

[18]The seemingly different results between the estimates for children who perform relatively well in the reading and math scores reported in panels A and B of Table 2 (see columns 70th-90th Perc) and for those ranked seventh or above reported in Table 4 can be explained by the limited correlation between the class-level rankings as implied by the average diagnostic test and the math (0.51) and reading scores (0.52).

4 Program Scale-Up and Persistence of the Impacts

Over the summer of 2016, after learning about the experimental impacts of the program, CONAFE decided to expand the API Plus model to all the eligible primary schools (see Appendix A.2), including the 224 schools that were part of the original evaluation sample. The Plus modality became the only modality in place for the API intervention thereafter.[19] In this section, we test whether the effects observed during the experimental evaluation persist at scale.

We observe the schools that were part of the evaluation sample up two years after the end of the experiment. By the spring of 2018, the years of exposure of schools to the API Plus range from zero to four depending on both the original experimental assignment in 2014 as well as on the 2016 and 2017 CONAFE reassignment. Approximately 60 percent of the schools in any of the original treatment arms received the API Plus program for at least one year during the national roll out.

One may be concerned that comparing children with respect to their total exposure to the intervention may confound differences in unobservable characteristics between schools. For this reason, we exploit the original treatment assignment in our experiment in 2014 to predict the school-level years of exposure to the API Plus program during the roll out. The schools that were originally assigned to the API Plus treatment can be exposed to the program from two to four years, while schools that were assigned to either the original control group or the API Standard group can be exposed to the API Plus program for up to two years (during the program roll out).[20]

We implement this research design using the following regression model

$$ExpPlus_j = \psi_0 + \psi_0 Plus_j + \psi_1 Standard_j + \psi_2' X_{i,j} + \nu_{i,j} \quad (2)$$

$$Y_{i,j} = \gamma_0 + \gamma_1 ExpPlus_j + \gamma_2 Standard_j + \gamma_3' X_{i,j} + \eta_{i,j}\ , \quad (3)$$

which we estimate via an instrumental variable (IV) estimator. The coefficient γ_1 is our parameter of interest, which represents the marginal effect on the outcome Y of increasing the API Plus exposure ($ExpPlus$) by one additional year.

In panel A of Table 5 we show that, on average, after one year of the program scale-up, children in the original API Plus schools are 18 percentage points more likely to enroll in secondary school relative to children in the original control schools. Within the same time horizon, children in the original API standard schools are 10 percentage points more likely

[19]Schools were assigned a score between 1 and 4, with 1 denoting the highest priority level. The scores are based on a combination of criteria that include school performance in the national learning assessment, whether the school had received an API during the period between 2008 and 2015, the level of marginalization of the community where the school was based, and whether the community was targeted by the anti poverty program *Cruzada Nacional contra el Hambre*.

[20]The first-stage estimates are reported in the Appendix Table E.10. They confirm that the original treatment assignments predict the total number of years of API exposure, with an average of approximately two additional years relative to the control group schools.

Table 5: Program Impacts at Scale

Panel A: Transition to Secondary School (Fall 2017)

	Reduced-Form	IV	OLS
Original API Standard	0.100	0.121	0.115
	(0.062)	(0.066)	(0.066)
Original API Plus	0.185		
	(0.066)		
Years of Exposure to API Plus		0.094	0.087
		(0.033)	(0.028)
Observations	625	625	625
Clusters	207	207	207

Panel B: School is Open (Fall 2018)

	Reduced-Form	IV	OLS
Original API Standard	0.133	0.148	0.151
	(0.057)	(0.057)	(0.053)
Original API Plus	0.152		
	(0.061)		
Years of Exposure to API Plus		0.076	0.080
		(0.029)	(0.021)
Observations	224	224	224

Notes: This table shows the estimates for the model in (2)–(3). In panel A we focus on the probability a student in 6th grade enrolling in a secondary school in the fall of 2017. In panel B we focus on the probability of a school in our original experiment being open by the fall of 2018. The first column shows the reduced-form estimates for both outcomes with respect to the original treatment assignments in our experiment. The second column shows the IV estimates for the years of total API exposure, when we use the original treatment assignments as excluded instruments. Finally, the third column shows the OLS estimates. Standard errors (in parentheses) are clustered at the school level in panel A.

to enroll in secondary school, but the effect is not statistically different from zero. Because both the original control schools and the original API standard schools started to participate in the API Plus program after the scale-up initiative, these reduced-form effects suggest—in line with the results shown in Table 1—that API Plus is the effective format.

By predicting years of exposure based on the initial randomization, the second-stage estimates in panel A of Table 5 show that an additional year of the API program increases the probability of children's transitioning into secondary school by 9.4 percentage points. The magnitude of this effect closely resembles the 8.8 percentage points experimental estimate of the API Plus displayed in the last column of Table 1.[21]

There are a few implementation details of the experiment that may threaten its scalability (see Appendix B). In particular, the threshold number of enrolled students to keep the school open was reduced from six to three as a way to minimize sample attrition during the experiment. However, the original requirement of six enrolled students was restored during the national roll out. For this reason, a threat to the success of the program at scale may come from the possible school closures following the API assignment. If instead the API assignment increases the probability of the school being open, this mechanism could bolster the external validity of the experiment.

We test this hypothesis using data on school closures through the follow-up survey conducted in the fall of 2018 (two years after the scale-up initiative and four years after the inception of the experiment). We present the results in panel B of Table 5. In this case, the reduced-form estimates show that schools in both the original API standard group and API plus group, after four years, are more likely to remain open in the fall of 2018 (+13 percentage points and +15 percentage points). This result suggests that the API assignment during the experiment created incentives for schools to remain open in the medium run. For instance, the parents' association can ultimately decide whether to keep the school in operation. Hence, one possible hypothesis behind these effects is that parental investments can generate spillover effects within a community, positively affecting the destiny of a school (List et al., 2019). The IV estimates imply that an additional year of API exposure leads to a 7.6 percentage points increase in the probability of a school's remaining open. This is a relatively large effect to the extent that 30 percent of the schools in the original control group were found to be closed by the fall of 2018.[22]

The OLS estimates of the regression model in (3) are remarkably similar to the IV estimates in both panels of Table 5, suggesting that the scale-up initiative did not select schools based on unobserved factors that determine our outcomes. We leverage this last result to provide additional evidence of the exposure effects of the program. We study the differential API exposure across schools after the program scale-up. Appendix Figure E.3 shows that for both transition and school closure outcomes (panels A and B), the estimated effects are linearly increasing in the years of exposure to the program. This result highlights the benefits of prolonged exposure for the program beneficiaries, which contrasts with the short-time span of the original design (two years). Finally, panels C and D of Appendix Figure E.3 document positive exposure effects of the API program on the reading and math test scores collected during the follow-up survey of the fall of 2018. However, these last results on achievement

[21] The main experimental estimate of API Plus is an average between the sixth graders in the 2015–2016 school year (one year of exposure), and the sixth graders in the 2015–2016 school year (two years of exposure). Splitting the sample by school years, we find that the API Plus increases the probability of transitioning to secondary school by 6.3 percentage points for an exposure of one year, and by 12.4 percentage points for two years.

[22] The importance of keeping schools open for the development of children has recently gained the attention of educational studies on the impact of the COVID-19 lockdowns on children in developed countries (see, e.g., Agostinelli et al., 2020; Engzell et al., 2020; Maldonado and De Witte, 2020). In the context of a developing country, Duflo (2001) finds that opening schools causes an increase in years of education for children.

scores are only suggestive due to the school closures over the same period (see panel B).[23]

5 Conclusion

We provide evidence on both experimental and at scale impacts of a mobile mentor program implemented in highly marginalized areas of Mexico. The program is particularly effective when augmented with an extra week of hands-on training on foundational skills and peer-to-peer sessions. Parental aspirations and investments seem to be the main channel through which the program affects children's outcomes.

This result provides new insights into the role of parents in shaping the success of education interventions targeted at school-age children, building upon the previous evidence on the key role of parents during early childhood (Heckman et al., 2010, 2013; Chaparro et al., 2020). More broadly, our findings suggest that a scalable intervention in a low-resource environment can effectively contribute to closing the gap between children in urban and rural communities.

References

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti, "When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times," 2020. NBER Working Paper 28264.

Al-Ubaydli, Omar, John A. List, and Dana Suskind, "The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments," NBER Working Papers 25848, National Bureau of Economic Research, Inc May 2019.

Albornoz, Facundo, María Victoria Anauati, Melina Furman, Mariana Luzuriaga, María Eugenia Podestá, and Inés Taylor, "Training to Teach Science: Experimental Evidence from Argentina," *World Bank Economic Review*, 2020, *34* (2), 393–417.

Andrew, Alison, Orazio P. Attanasio, Britta Augsburg, Monimalika Day, Sally Grantham-McGregor, Costas Meghir, Fardina Mehrin, Smriti Pahwa, and Marta Rubio-Codina, "Effects of a scalable home-visiting intervention on child development in slums of urban India: evidence from a randomised controlled trial," *Journal of Child Psychology and Psychiatry*, 2020, *61* (6), 644–652.

Attanasio, Orazio P., Camila Fernández, Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, and Marta Rubio-Codina, "Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early

[23]Only two schools were permanently closed at the beginning of the program roll out in the school year 2016–2017. Therefore, the sample of sixth graders for which we observe the transitions into secondary school (see panel A of Table 5) is not affected by selective attrition.

child development program in Colombia: cluster randomized controlled trial," *BMJ*, 2014, *349.*

— , **Costas Meghir, and Emily Nix**, "Human Capital Development and Parental Investment in India," *The Review of Economic Studies*, 2020, *87* (6), 2511–2541.

— , **Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina**, "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia," *American Economic Review*, 2020, *110* (1), 48–85.

Bando, Rosangela and Claudia Uribe, "Experimental Evidence on Credit Constraints," Working Paper 670, Inter-American Development Bank February 2016.

Banerjee, Abhijit V., Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton, "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, November 2017, *31* (4), 73–102.

— , **Shawn Cole, Esther Duflo, and Leigh Linden**, "Remedying Education: Evidence from Two Randomized Experiments in India*," *The Quarterly Journal of Economics*, 2007, *122* (3), 1235–1264.

Bassi, Marina, Costas Meghir, and Ana Reynoso, "Education Quality and Teaching Practices," *The Economic Journal*, 05 2020, *130* (631), 1937–1965.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur, "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*, 2018, *168* (C), 1–20.

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

Cameron, Lisa, Susan Olivia, and Manisha Shah, "Scaling up sanitation: Evidence from an RCT in Indonesia," *Journal of Development Economics*, 2019, *138*, 1–16.

Carneiro, Pedro, Emanuela Galasso, Italo Lopez Garcia, Paula Bedregal, and Miguel Cordero, "Parental Beliefs, Investments, and Child Development: Evidence from a Large-Scale Experiment," IZA Discussion Papers 12506, Institute of Labor Economics (IZA) July 2019.

— , **Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott**, "The Impacts of a Multifaceted Pre-natal Intervention on Human Capital Accumulation in Early Life," IZA Discussion Papers 13955, Institute of Labor Economics (IZA) December 2020.

Chaparro, Juan, Aaron Sojourner, and Matthew J. Wiswall, "Early Childhood Care and Cognitive Development," NBER Working Papers 26813, National Bureau of Economic Research, Inc February 2020.

Duflo, Esther, "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *The American Economic Review*, 2001, *91* (4), 795–813.

_ **, Rema Hanna, and Stephen P. Ryan**, "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, June 2012, *102* (4), 1241–1278.

Engzell, Per, Arun Frey, and Mark Verhagen, "Learning Inequality During the Covid-19 Pandemic," 2020. Mimeo.

Fryer, Roland G., Steven D. Levitt, and John A. List, "Parental Incentives and Early Childhood Achievement: A Field Experiment in Chicago Heights," NBER Working Papers 21477, National Bureau of Economic Research, Inc 2015.

Gertler, Paul, James Heckman, Rodrigo Pinto, Arianna Zanolini, Christel Vermeersch, Susan Walker, Susan M. Chang, and Sally Grantham-McGregor, "Labor market returns to an early childhood stimulation intervention in Jamaica," *Science*, 2014, *344* (6187), 998–1001.

Heckman, James J. and Stefano Mosso, "The Economics of Human Development and Social Mobility," *Annual Review of Economics*, 2014, *6*, 689–733.

_ **, Rodrigo Pinto, and Peter Savelyev**, "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes," *American Economic Review*, October 2013, *103* (6), 2052–86.

_ **, Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz**, "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 2010, *94* (1), 114–128.

List, John A., Fatemeh Momeni, and Yves Zenou, "Are Estimates of Early Education Programs Too Pessimistic? Evidence from a Large-Scale Field Experiment that Causally Measures Neighbor Effects," 2019. CEPR Discussion Paper 13725.

Maldonado, Joana E. and Kristof De Witte, "The Effect of School Closures on Standardised Student Test Outcomes," 2020. KU Leuven Discussion Paper Series 20.17.

Muralidharan, Karthik and Abhijeet Singh, "Improving Public Sector Management at Scale? Experimental Evidence on School Governance India," Working Paper 28129, National Bureau of Economic Research November 2020.

_ **and Paul Niehaus**, "Experimentation at Scale," *Journal of Economic Perspectives*, 2017, *31* (4), 103–124.

Ozler, Berk, Lia C. H. Fernald, Patricia Karol Kariger, Christin Mcconnell, Michelle J. Neuman, and Eduardo Pinheiro Fraga, "Combining preschool teacher training with parenting education : a cluster-randomized controlled trial," Policy Research Working Paper Series 7817, The World Bank September 2016.

Parente, Paulo M.D.C. and Joao M.C. Santos Silva, "Quantile Regression with Clustered Data," *Journal of Econometric Methods*, 2016, *5*, 1–15.

World Bank, "World Development Report 2018 : Learning to Realize Education's Promise," Technical Report, Washington, DC: World Bank 2018.

Yoshikawa, Hiroshi, Diana Leyva, Catherine Snow, Trevino E. Ernesto, Clara M. Barata, Cristina Weiland, and Mary C. Arbour, "Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes," *Developmental psychology*, March 2015, *51* (3), 309—322.

Appendices

A Institutional Context

A.1 The CONAFE Multi-Grade System

In the year 2014, CONAFE schools accounted for 10 percent of the roughly 99,000 primary schools and 7 percent of the 38,000 lower secondary schools in Mexico.

CONAFE instructors are generally young community residents between 15 and 29 years old who have completed the upper secondary education cycle. They do not have formal teacher training and they are supposed to stay in the same school for two years. Only 2.6 percent of CONAFE teachers report having a college degree, while 19 percent report having only completed lower secondary education. Instructors should receive between five and seven weeks of training, but more than half report four weeks of training or less. They receive a stipend of MXN $1,427 per month (USD $95 in 2015). After one year of service, instructors become eligible to receive a scholarship of MXN $982 per month for up to 30 months, which is conditional on enrolling in a higher education institution. As a result of the very low compensation and extremely challenging conditions, about one quarter of the instructors drop out before completing the school year.

Communities that receive CONAFE services organize a local association aimed at promoting community education, which is responsible for providing instructors with the accommodations, meals, and security they need to reside in the community (*Diario Oficial de la Federación* 2012). In principle schools that have more than 29 students are eligible to become part of the regular system, but it is often the case that communities prefer to maintain CONAFE status.

In the 2012–2013 school year, on average 23 percent of the instructors quit before the end of the school year. Qualitative evidence suggests that the low salary and the difficult conditions in the community are the most common reasons for quitting the job. According to Bando and Uribe (2016), 62 percent of the instructors reported that the local association was not organized to provide food and lodging when they arrived, and 46 percent reported having slept in the school, and 62 percent said they spent money on food.

A.2 Additional Details on the API Intervention

CONAFE schools receive the API program in pairs, with the mentors spending an equal amount of time in both communities (see also Section 2.1 of the main draft). Each pair consists of one school, which is the main target of the intervention, and another school that is included on the basis of proximity to the target school. Target schools are selected according to the following criteria: (i) they have at least 30 percent of the students classified as "insufficient" in the Nationwide Standardized test ENLACE; and (ii) they have at least six students enrolled in primary school. Among the schools that met the above criteria, preference was given to the municipalities with communities that are characterized by high levels of poverty (as proxied by the presence of the anti poverty program, the National Crusade Against Hunger), difficulty of access, and a large presence of indigenous communities. Both target and auxiliary schools can, in principle, refuse the assignment of a mentor, which may happen if, for instance, the local communities lack resources to satisfy the mentors' basic needs (lodging and meals).

In 2013 there were 2,099 mentors operating in CONAFE schools across the 31 Mexican states. Dropout among mentors is higher than among community instructors (about 40 percent per school year), as they are more likely to find jobs that are either better paid or are based in better locations. Indeed, it is quite common for the mentors to switch to a regular teacher job, as a result of periodic national recruitment drives.

Student eligibility for the remedial education sessions is determined by a joint assessment of the instructor and the mentor, which is based on a diagnostic evaluation that the instructor conducts at the beginning of the school year as well as the student's difficulties in reading and basic math and having repeated one or more grades. The evaluation covers the material that students should have mastered in the previous grade. It covers seven subjects and the grades vary in between 5 and 10. Preference is given to students in the 3rd to 6th grades. Once the eligible students have been identified, the mentor administers an additional exam, the *examen de colocación*, to establish the effective grade to which the student's knowledge corresponds. With this information in hand, the mentor prepares a personalized plan. Throughout the school cycle, the mentor provides the students with constant feedback and constantly monitors their progress through a personalized evaluation form, the *Cuadernillo para el Seguimiento del Alumno*.

B Additional Details on the Evaluation

Some of the insights for the design of the evaluation were drawn from a survey collected in 2011 from 40 communities over four Mexican states as part of a World Bank project. About two thirds of the mentors did not speak any indigenous language, although a large group of students reported an indigenous language as the only language spoken. In addition, only one-third of the parents reported that the mentor had spent one hour or less doing home visits. Finally, one-third of the mentors reported that the training module was not helpful in addressing the pedagogical challenges they were facing in the communities.

The API Standard modality tracks the benchmark interventions with two minor differences. First, the ability to speak the main indigenous language in the community would become the most important criterion for the assignment of the mobile mentors across program-eligible communities. Second, the supervisors of the mobile mentors would receive a salary increase in exchange for a mandatory increase in the frequency of their visits to the targeted communities.

While the official enrollment threshold for closing CONAFE schools is six students, the schools in our sample were allowed to remain open if they had at least three enrolled students in either school years 2014–2015 and 2015–2016. In addition, evaluation schools with more than 29 enrolled were not required to transition into the regular public-school system. To avoid refusal of the assigned mentor among the communities of the target schools, each mentor in the evaluation sample would donate two food baskets to the community leaders (one at the beginning and the other at the end of the school year 2014–2015). This strategy was indeed effective as no schools in the evaluation sample refused the assigned mentors. Similar arrangements could not be put in place in the context of auxiliary schools that were matched with the target schools in the evaluation sample. More broadly, limited financial resources did not allow either to monitor the implementation of the intervention or to collect survey data in auxiliary schools.

As a way to attenuate the potentially detrimental consequences of mentors' dropping out of the program during the evaluation period, the CONAFE delegation in Chiapas arranged for a replacement within two weeks from the day of the departure from a community. The replacement candidate would be randomly selected from those who had served during the previous school cycle and who complied with the new eligibility criteria outlined in Section 2.1. If the dropout was part of the Plus group, the replacement would receive an additional three-day training session that would make up for the content covered during the extra week of the initial training session.

C Data Description

C.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. In the case of CONAFE schools, 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 panel A of Appendix Table E.1. Census data for the years 2016 and 2017 are used to track the schools in the evaluation sample during the scale-up.

Standardized test scores. Between 2007 and 2013, all Mexican students in grades 3 through 9 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 and geography. We use the school-level average of the 2012 data to construct the strata for the school-level randomization and school-level averages for 2013 as controls in all our regression models.

CONAFE records. All students in the CONAFE system, irrespective of whether they receive the API program or not, have to undergo a diagnostic exam at the beginning of each school year. The exam covers three subjects: math, Spanish, and natural science. The score for each subjects ranges between 5 and 10. We use the individual level average across the three subjects in the diagnostic exams at the beginning of the school year 2014–2015 to construct the student ranking within each CONAFE school displayed in Figure E.2. From that variable, we construct the indicator variables $RankAbove7$ and $RankBelow7$ reported in Table 4, which proxy for the individual eligibility to the one-on-one remedial education sessions.

CONAFE also collects information about their instructors and mobile mentors, such as age, gender, education attained, overall experience and tenure in the school. These baseline characteristics are reported in panels B and E of Table E.1.

Transition rates from primary to lower secondary We link the enrollment records of the sixth graders in our sample across the first grade of secondary education (7th grade) both in the CONAFE and the regular public education (SEP) system during the following academic year. Individual transitions computed in the school years 2015–2016, 2016–2017 are reported in Tables 1 and 2, while transitions computed in the school year 2017–2018 are reported in Table 5 and in Figure E.3.

C.2 First follow-up survey in 2016

The first round of data collection took place in the spring of 2016 in the 224 schools that form part of the evaluation sample. It entails the following array of survey modules and measurement tools.

Measures of Cognitive Achievement. The Early Grade Reading Assessment (reading score) and the Early Grade Math Assessment (math score) are individually administered oral student assessments that have been conducted in more than 40 countries and in a variety of languages. While these instruments are typically applied to students in first, second, or third grade, we administer them to all the students enrolled in 3rd to 6th grades in the 224 schools of the evaluation sample to account for the large learning gaps of these children when compared to the national average (see Section 2.2).

The reading score reported in Tables 1, 2, 4, and 5 is 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) listing, 8) dictation. The math score reported in Tables 1, 2, 4, and 5 is 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. The individual components of the math and reading score are reported in Table E.2. An orthogonal rotation is applied before standardizing each factor with respect to the mean and the standard deviation in the control group.

Measures of Socio-Emotional Development. The household survey contains a set of measures of behavioral problems reported by the caregivers of the children in our sample. The socio-emotional score reported in Tables 1, 2, 4 is the sum of the following thirty two items on how often the child displays a given emotion/behavior (1 = Never, 2 = Sometimes, 3 = Always).

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 scary or anxious
6. Talks and argues too much
7. Has difficulties in focusing on a specific activity for an extended amount of time
8. Gets easily confused. It seems that his/her head is in the clouds
9. Threatens or is mean with other children
10. Tends to challenge parental authority
11. Does not feel guilty after a bad dead
12. Does not get along with other children
13. Is impulsive or act "fast" without thinking
14. Feels has inferiority issues
15. Has no friends
16. Has difficulties to let go certain thoughts
17. Is hyper-active
18. Has a bad temper, or is irascible
19. Looses easily his/her temper
20. Feels unhappy, sad, or depressed
21. Is shy, does not socialize with others
22. Breaks objects on purpose
23. Is too attached to the adults

24. Cries too much
25. Demands a lot of attention
26. Is too much dependent on others
27. Afraid of other people's judgement
28. Tends to be in bad company
29. Reserved, keeps things for himself/herself
30. Worries about every thing
31. Misbehaves at school
32. Does not respect the instructor

We sum the scores of the thirty two items and we re-scale the resulting index in such a way that higher values are associated with fewer behavioral issues.

Household Survey. Since surveying the universe of households was not feasible given the budget of the evaluation, a random sample of five households was selected within a 5 kilometer radius from each school in our sample. Basic information on both the household module respondent and household characteristics is reported in panel D of Table E.1.

The household module also collected information on parents' expectations and investment toward children's education as well as measures of homework supervision, interactions with community instructors, time spent on a number of school-related activities, and number of books at home. The parental expectation variable reported in panel A of Table 3 takes the value of one if the respondent expects her/his child to complete upper secondary education or higher, and zero otherwise. The parental investment index also reported in panel A of Table 3 is given by the latent factor of an exploratory factor analysis of five components: (i) an indicator variable for whether the caregiver helps with homework, (ii) participation in school fund-raising activities (school activities 1), (iii) participation in other school-related activities (school activities 2), (iv) the frequency of meetings with the child's instructor, and (v) whether the child participates in any academically-related activities outside the school hours. The individual components are reported in Table E.7. The factor is obtained with Bartlett's method, where the factor scores highly correlate with its own factor and not with others.

Finally, the household module collected 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 Standard group or the API Plus group. This information is used to construct the four variables reported in panel B of Table 3.

Measures of Pedagogical Practices and Instructors characteristics. 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—a rubric for timed observations that has been previously used 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 or not. 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 is used for the two outcome variables of the community instructors used in the last two columns of panel A in Table 3. The teaching score reported in Table E.6 is constructed as the sum of the individual scores in each three-minute snapshot for five key aspects of pedagogy: (i) reading, (ii) showing, (iii) answering questions, (iv) memorizing, and (v) giving homework. The material score also reported in Table E.6 is constructed as 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 exposed in the classroom, and (iv) whether charts and maps are exposed in the classroom.

Instructors were also asked standard questions on their socio-demographic characteristics, education, experience and, if they were in the treatment group, their relationship with the mentors. Those are reported in panel B of Table E.1.

Mentors' Survey. Since the pedagogical mentors were not located in the communities on a continuous basis, the survey firm interviewed them by an end-of-year evaluation session. Their characteristics are reported in Table E.3 for the subset of the mentors who reported working in different schools from those they were initially assigned to.

C.3 Second follow-up survey in 2018

In the fall of 2018, we conducted a second round of data collection to evaluate the medium-term impacts of the intervention. As mentioned in Section 4, a significant number of schools in the original evaluation sample were closed during the time between the two data collections. Children's outcomes were measured in 186 schools out of the 224 original schools that were part of the evaluation. Using this information, we have computed the indicator function for whether the school is open that we used in Table 5 and Figure E.3. Both the reading score and the math score (see Section C.2) were measured for children in grades three through six and they are reported in Figure E.3.

D Standard versus Plus: The Role of the Remedial Education Sessions

In this section, we provide more details on the interpretation of the results reported in Table 4. The goal is to decompose the heterogeneity of the effects above and below rank 7 into: (i) the direct effect of the two of API modalities for low-achieving and high-achieving children; and (ii) the effect of the remedial education sessions on children. For this reason, we first allow the direct effect of the API Standard and API Plus modalities to vary depending upon the children's rank in the class ($Below7_i$), independently of the remedial sessions. Second, we allow the effect of the remedial education sessions ($Remedial_i$) to vary between the two API modalities.

The full-interacted model takes the following form:

$$\text{(D.1)} \quad \begin{aligned} Y_{i,j} =& \beta_{0,0}(1 - Below7_i) + \beta_{0,1} Below7_i + \\ &+ \beta_{1,0}(1 - Below7_i) * Standard_j + \beta_{2,0}(1 - Below7_i) * Plus_j + \\ &+ \beta_{1,1} Below7_i * Standard_j + \beta_{2,1} Below7_i * Plus_j + \\ &+ \beta_3 Remedial_i * Standard_j + \beta_4 Remedial_i * Plus_j + u_{i,j} \,, \end{aligned}$$

where the coefficients $(\beta_{1,0}, \beta_{2,0})$ and $(\beta_{1,1}, \beta_{2,1})$ represent the effect of API Standard and API Plus for children above and below rank 7, respectively. The coefficients (β_3, β_4) represent the effect of the remedial education sessions for the API Standard and the API Plus, respectively. These latter effects cannot be identified in our context as we don't have variation in $Remedial_i$ for the control group and the assignment to the remedial sessions was not randomized within the Standard and Plus groups. Still, we can decompose the differential impact of API Plus versus API Standard between children above or below rank 7 as follows:

$$\text{(D.2)} \quad \begin{aligned} &(E[Y|Plus = 1, Below7 = 1] - E[Y|Standard = 1, Below7 = 1]) - \\ &(E[Y|Plus = 1, Below7 = 0] - E[Y|Standard = 1, Below7 = 0]) = \\ &\underbrace{\underbrace{(\beta_{2,1} - \beta_{1,1})}_{\substack{\text{Relative Effect of Plus vs Standard} \\ \text{Below Rank 7}}} - \underbrace{(\beta_{2,0} - \beta_{1,0})}_{\substack{\text{Relative Effect of Plus vs Standard} \\ \text{Above Rank 7}}}}_{\geq 0} + \underbrace{\underbrace{(\beta_4 - \beta_3)}_{\substack{\text{Contribution of Remedial Sessions on} \\ \text{Plus vs Standard Effects}}}}_{\leq 0} \approx 0 \,, \end{aligned}$$

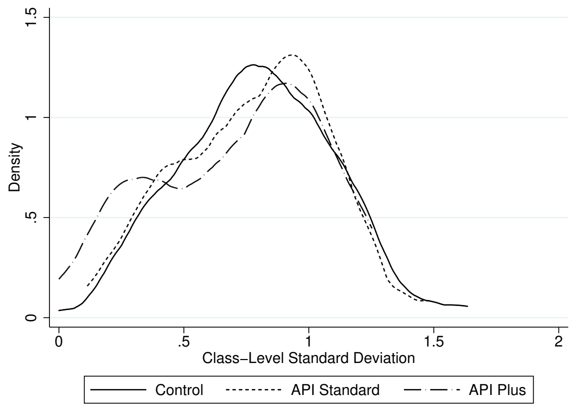
The main object of interest in this expression is $\beta_4 - \beta_3$, which represents the unobserved contribution of the remedial education sessions in explaining the observe difference in education outcomes between the effects of API Plus and API standard.

The estimates reported in Table 2 document that the differential impact of API Plus versus API standard decreases over the quantiles of the distribution of reading and math scores. Hence, we can bound the difference between the first two terms of equation (D.2) to be positive ($(\beta_{2,1} - \beta_{1,1}) - (\beta_{2,0} - \beta_{1,0}) \geq 0$). The overall term in (D.2) is assumed to be zero because of the estimated coefficients reported in Table 4. This implies that the remedial educational sessions cannot explain the differential impacts between the API Plus and the

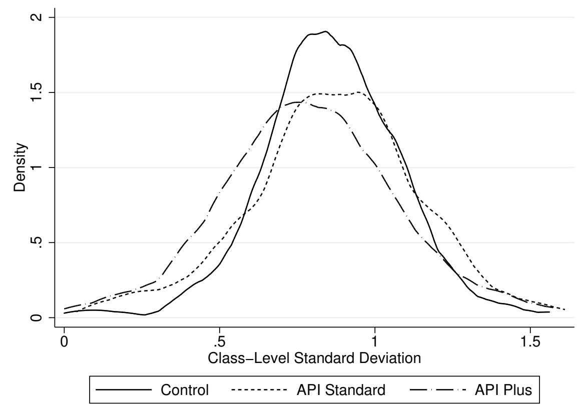
API Standard on a child's development ($\beta_4 - \beta_3 \leq 0$).

E Additional Figures and Tables

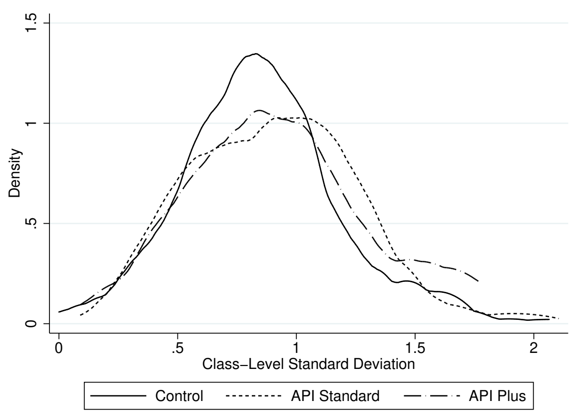
Figure E.1: Program Impacts on the Class-Level Dispersion of Learning Outcomes



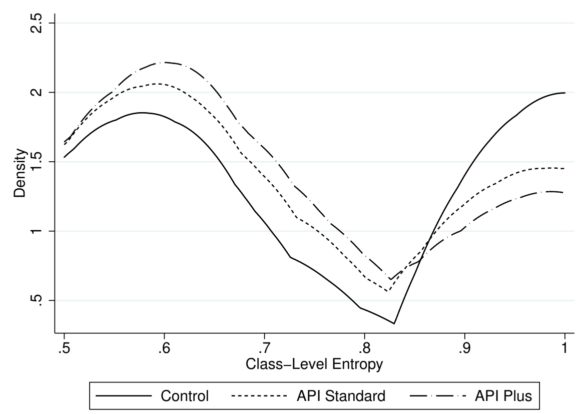
Panel A: Reading Score



Panel B: Math Score



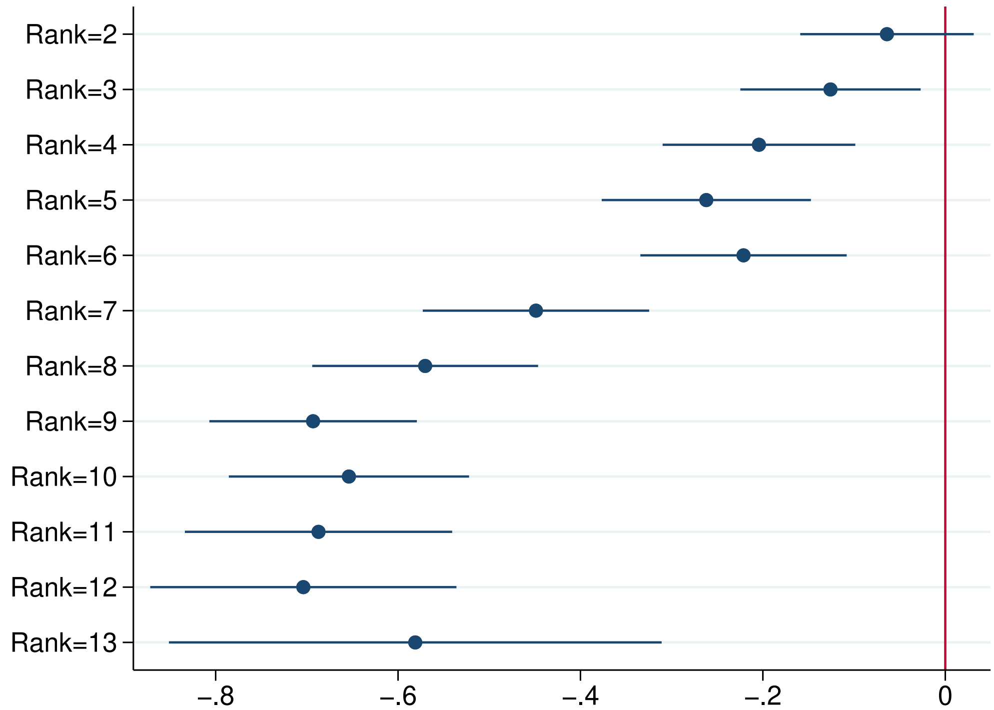
Panel C: Socio-Emotional Score



Panel D: Transition to Secondary

Notes: This figure shows the distribution of the class-level dispersion in outcomes between the three treatment arms. For each school in our sample we calculate either the standard deviation (reading, math, and socio-emotional scores) or the entropy index (transition to secondary) for the various outcomes.

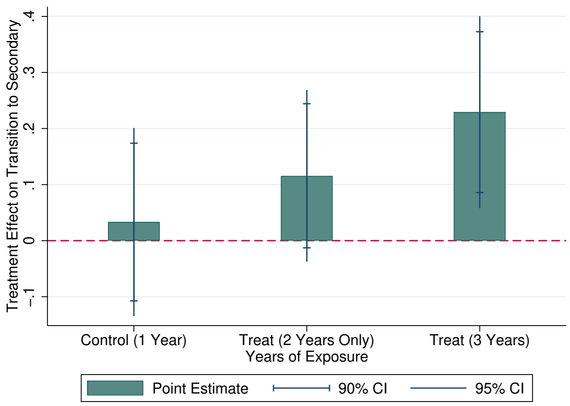
Figure E.2: Marginal Effects on the Probability of Being in the Remedial Sessions by Inverted Achievement Ranks



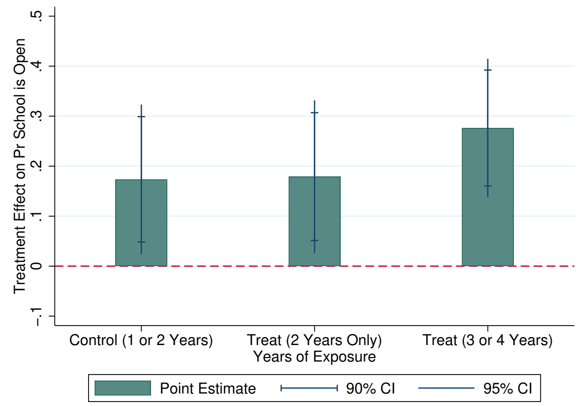
Notes: This figure shows the differential probability of students being assigned to the one-on-one remedial sessions as a function of their within-class relative rank in the diagnostic test. The omitted category is Rank = 1, which represents the least-performing child in the class.

Figure E.3: The Effects of the Length of Program Exposure

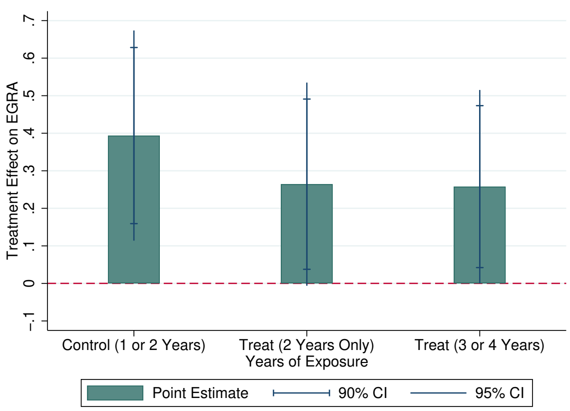
Panel A: Transition to Secondary School



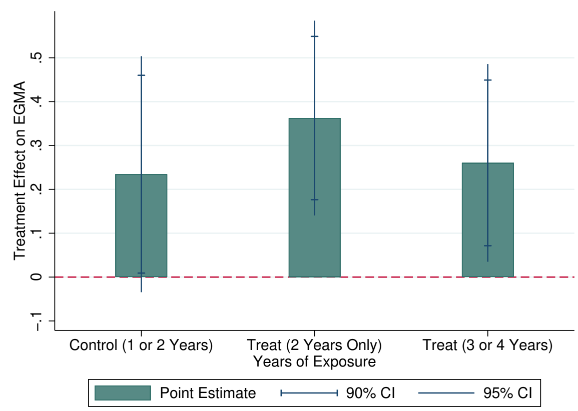
Panel B: Probability of School Being Open



Panel C: Reading Score



Panel D: Math Score



Notes: The figure shows the differential effects of API exposure through the program national scale-up on the probability of transition to secondary school in the fall of 2017 (panel A), on the probability of the schools to remain open in the fall of 2018 (panel B), as well as on reading and math scores (panel C and D). We divide the schools based on the API exposure from 2014 to 2018 into four categories: schools in the original control group with no exposure (omitted category), schools in the control group with 1 or 2 years of exposure, schools in either one of the two treatment arms with 2 years of exposure, and school in the original treatment arms with 3 or more years of exposure.

Table E.1: Covariate Balance

Variable	Control Mean (SD)	API Standard Mean (SD)	API Plus Mean (SD)	Difference Std-Ctr (SE)	Difference Plus-Ctr (SE)
Panel A: Schools Characteristics					
ENLACE Spanish 2012	431.878	431.654	431.361	-0.223	-0.516
	(64.431)	(66.869)	(66.604)	(2.581)	(2.783)
ENLACE Math 2012	455.754	454.852	451.678	-0.902	-4.076
	(80.710)	(83.505)	(81.063)	(5.790)	(6.911)
ENLACE Science 2012	440.149	441.244	441.269	1.095	1.120
	(52.523)	(48.665)	(50.895)	(4.273)	(4.784)
Community Instructors in 2013	1.220	1.300	1.200	0.080	-0.020
	(0.416)	(0.462)	(0.403)	(0.066)	(0.067)
Enrollment in 2013	15.160	15.314	14.233	0.154	-0.927
	(5.839)	(5.714)	(5.782)	(0.901)	(0.946)
Panel B: Community Instructors Characteristics					
Lower than upper second.	0.067	0.062	0.066	-0.002	0.009
	(0.251)	(0.242)	(0.250)	(0.035)	(0.033)
Lower than higher ed.	0.918	0.901	0.908	-0.000	0.002
	(0.276)	(0.300)	(0.291)	(0.044)	(0.040)
Training weeks at baseline	4.515	4.704	4.500	0.128	-0.042
	(1.322)	(1.259)	(1.426)	(0.196)	(0.253)
3rd and 4th grade students	3.655	3.986	3.716	0.346	0.137
	(2.434)	(2.286)	(2.230)	(0.349)	(0.356)
5th and 6th grade students	3.517	3.838	3.507	0.325	0.054
	(2.408)	(2.507)	(2.298)	(0.354)	(0.352)
Panel C: Student Characteristics					
Baseline Age (Months)	104.993	104.289	105.539	-0.818	0.647
	(16.384)	(17.532)	(14.924)	(1.169)	(1.247)
Male	0.533	0.519	0.543	-0.012	0.011
	(0.499)	(0.500)	(0.499)	(0.033)	(0.043)
Spanish Test score	7.499	7.634	7.546	0.143	0.043
	(0.887)	(0.829)	(0.778)	(0.100)	(0.089)
Math Test score	7.632	7.720	7.617	0.096	-0.018
	(0.931)	(0.842)	(0.833)	(0.100)	(0.092)
Natural Science Test score	7.579	7.710	7.582	0.137	-0.003
	(0.856)	(0.773)	(0.731)	(0.097)	(0.092)
Social Sciences Test Score	7.416	7.585	7.451	0.174	0.031
	(0.859)	(0.782)	(0.750)	(0.102)	(0.097)
Repeater	0.024	0.003	0.012	-0.022	-0.013
	(0.152)	(0.057)	(0.107)	(0.011)	(0.012)
Panel D: Household Characteristics					
Indigenous Language	0.326	0.366	0.476	0.049	0.142
	(0.469)	(0.483)	(0.500)	(0.065)	(0.077)
Read	0.715	0.686	0.734	-0.031	0.022
	(0.452)	(0.465)	(0.443)	(0.041)	(0.042)
Less than Primary	0.615	0.587	0.584	-0.028	-0.030
	(0.487)	(0.493)	(0.494)	(0.043)	(0.041)
Upper Sec. or Higher	0.015	0.016	0.019	-0.001	0.003
	(0.123)	(0.124)	(0.135)	(0.009)	(0.009)
Oportunidades	0.813	0.807	0.829	-0.003	0.015
	(0.391)	(0.395)	(0.377)	(0.033)	(0.031)
Refrigerator	0.397	0.387	0.373	-0.010	-0.019
	(0.490)	(0.488)	(0.485)	(0.047)	(0.055)
Television	0.692	0.738	0.651	0.048	-0.040
	(0.462)	(0.440)	(0.478)	(0.047)	(0.051)
Car	0.084	0.081	0.063	-0.003	-0.019
	(0.277)	(0.273)	(0.244)	(0.027)	(0.024)
Sewage	0.254	0.253	0.320	-0.003	0.068
	(0.436)	(0.435)	(0.467)	(0.042)	(0.052)
Phone	0.220	0.233	0.204	0.014	-0.014
	(0.414)	(0.423)	(0.404)	(0.037)	(0.038)
Light	0.863	0.916	0.873	0.054	0.006
	(0.344)	(0.278)	(0.333)	(0.040)	(0.040)

Panel E: API Characteristics

Variable	API Standard Mean (SD)	API Plus Mean (SD)	Difference Plus-Std (SE)
Age	28.491	28.543	-0.135
	(3.760)	(3.075)	(0.650)
Male	0.566	0.587	-0.064
	(0.500)	(0.498)	(0.097)
High Edu Complete	0.887	0.891	0.014
	(0.320)	(0.315)	(0.066)
Previously Instructor	0.792	0.848	-0.079
	(0.409)	(0.363)	(0.072)
Previously Education Assistant	0.075	0.065	0.014
	(0.267)	(0.250)	(0.049)

Notes: School characteristics are based on the 2012 ENLACE and the 2013 Formato 911. Information on the characteristics of the community instructors draws on the 2016 follow-up survey, except for the number of training weeks that are reported in 2014 CONAFE administrative data. Student level information is based on the 2016 follow-up survey and the 2014 CONAFE administrative information on student diagnostics. Household level characteristics are obtained from the 2016 follow-up survey. The differences reported in the last two columns of the table are based on OLS regressions that control for stratification dummies. Standard errors for community instructors, student and household characteristics are clustered at school level.

Table E.2: Analysis of Reading and Math Scores by subdomain

Panel A: Share of Correct Math Answers by Sub-Domain

	Letter Name	Initial Name	Initial Sound	Word Recognition	Word Reading	Reading Comprehension	Listing	Dictation
API Standard	0.103	0.006	0.122	0.129	0.075	0.118	-0.004	0.129
	(0.086)	(0.079)	(0.085)	(0.076)	(0.072)	(0.073)	(0.078)	(0.083)
API Plus	0.240	-0.019	0.042	0.318	0.197	0.321	0.123	0.378
	(0.084)	(0.082)	(0.073)	(0.081)	(0.079)	(0.084)	(0.084)	(0.076)
H0: Standard=Plus	0.180	0.771	0.343	0.039	0.183	0.023	0.094	0.005
Mean Control Group	-0.000	-0.000	0.000	-0.000	-0.000	0.000	-0.000	-0.000
SD Control Group	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Observations	1044	1044	1044	1044	1044	1044	1044	1044
Clusters	224	224	224	224	224	224	224	224

Panel B: Share of Correct Reading Answers by Sub-Domain

	Num Identif	Num Discrim	Missing Num	Add	Subtract	Prob Solving	Shape Recog
API Standard	0.094	0.036	0.099	0.011	0.061	-0.051	0.022
	(0.082)	(0.081)	(0.076)	(0.068)	(0.072)	(0.073)	(0.081)
API Plus	0.259	0.201	0.204	0.215	0.111	0.116	0.099
	(0.091)	(0.089)	(0.089)	(0.072)	(0.068)	(0.082)	(0.098)
H0: Standard=Plus	0.095	0.103	0.218	0.008	0.500	0.046	0.396
Mean Control Group	-0.000	0.000	0.000	0.000	0.000	-0.000	0.000
SD Control Group	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Observations	1044	1044	1044	1044	1044	1044	1044
Clusters	224	224	224	224	224	224	224

Notes: This table shows the estimates of the two API modalities on each subdomain of Reading and Math scores. Each coefficient represents the estimated effect of the program relative to the control group (see the regression model in equation (1)). Standard errors in parentheses are clustered at the school level.

Table E.3: Characteristics of Dropout mentors

	Standard	Plus	Plus - Standard
Former CONAFE facilitator	0.689	0.703	0.012
	(0.468)	(0.463)	(0.102)
At least 5 days of training	0.467	0.514	0.061
	(0.505)	(0.507)	(0.111)
Sleeps in community (y/n)	0.711	0.757	0.052
	(0.458)	(0.435)	(0.097)
Number of nights in community last week	3.022	2.757	-0.301
	(2.061)	(1.978)	(0.442)
Number of students with personalized attention	6.049	5.767	-0.284
	(0.835)	(1.104)	(0.264)
Days spent in community during last month	10.220	10.200	0.063
	(4.613)	(4.715)	(1.148)
Insufficient students Level 2 according instructor	3.450	3.560	0.079
	(1.679)	(1.660)	(0.440)
Insufficient students Level 3 according instructor	2.727	2.731	-0.020
	(1.773)	(1.845)	(0.488)

Notes: This table shows the characteristics of the mentors (see Appendix C.2) who dropped out from the schools where they were originally assigned across API Standard and API Plus modalities.

Table E.4: API and Educational Inequality

	Reading Score Class-Level SD	Math Score Class-Level SD	Socio-emotional Class-Level SD	Transition to Secondary Class-Level Entropy
API Standard	-0.045	-0.004	0.059	-0.046
	(0.047)	(0.040)	(0.060)	(0.035)
API Plus	-0.131	-0.098	0.118	-0.059
	(0.053)	(0.045)	(0.069)	(0.034)
H0: Standard=Plus	0.126	0.067	0.430	0.746
Mean Dep. Var.	0.780	0.839	0.817	0.776
SD Dep. Var.	0.329	0.260	0.368	0.216
Observations	220	220	216	222

Notes: This table shows the estimates of the two API modalities on the class-level dispersion in outcomes. For each school in our sample we calculate either the standard deviation (Reading, Math and Socio-emotional scores) or the Entropy index (Transition to Secondary) for the various outcomes.

Table E.5: Treatment Assignment and Classroom Composition

	Has to repeat	Outside CONAFE system in previous A.Y	Same school as in previous A.Y.
API Standard	-0.011	-0.002	0.006
	(0.007)	(0.014)	(0.020)
API Plus	-0.010	-0.003	0.009
	(0.007)	(0.016)	(0.019)
H0: Standard=Plus	0.834	0.911	0.897
Mean	0.011	0.045	0.932
Observations	1019	1019	1045
clusters	224	224	224

Notes: This table shows the estimates of the two API modalities on various measures of changes in classroom composition: grade retention, inflows of new students from CONAFE, and inflows of students from outside the CONAFE system.

Table E.6: Pedagogical Practices

	Teaching Score	Material Score
API Standard	0.012	-0.170
	(0.136)	(0.149)
API Plus	-0.002	0.014
	(0.154)	(0.158)
H0: Standard=Plus	0.933	0.247
Mean Dep. Var.	-0.005	0.056
Observations	259	259
Clusters	209	209

Notes: This table shows the estimates of the two API modalities on two measures of pedagogical practices: teaching score and material score. For a detailed descriptions of the outcome variables used in this table, see Appendix C.2.

Table E.7: Parental Behaviors

	Helps with homework	School Activities 1	School Activities 2	Meeting with Teacher	Child Engaged in Extra-Activities
API Standard	0.051	-0.123	-0.168	0.046	0.072
	(0.034)	(0.085)	(0.097)	(0.075)	(0.042)
API Plus	0.082	0.152	0.080	0.195	0.112
	(0.042)	(0.087)	(0.097)	(0.081)	(0.045)
H0: Standard=Plus	0.464	0.007	0.036	0.113	0.409
Mean Control Group	0.479	-0.001	0.000	-0.006	0.636
SD Control Group	0.500	1.001	1.001	0.995	0.482
Observations	1043	1044	1044	973	1032
Clusters	224	224	224	223	224

Notes: This table shows the estimates for the behavioural responses in parental investments. The omitted category refers to the control group. For a detailed descriptions of the outcome variables used in this table, see Appendix C.2. Standard errors in parentheses are clustered at the school level.

Table E.8: Covariate Balance for Sub-Sample Below 7

Variable	(1) Control Mean (SE)	(2) API Standard Mean (SE)	(3) API Plus Mean (SE)	(4) Difference (2)-(1)	(3)-(1)
Panel A: Student					
Baseline Age (Months)	104.308	104.479	104.769	0.071	0.925
	(17.684)	(18.548)	(16.672)	(2.555)	(2.393)
Male	0.462	0.574	0.462	0.121	0.018
	(0.501)	(0.497)	(0.502)	(0.068)	(0.074)
Scholarship	0.778	0.755	0.782	-0.014	0.028
	(0.418)	(0.432)	(0.416)	(0.068)	(0.058)
Score Baseline Spanish. Test Conafe	8.085	8.019	8.096	-0.027	0.023
	(0.728)	(0.640)	(0.641)	(0.128)	(0.124)
Score Baseline Math Test Conafe	8.248	8.135	8.238	-0.084	-0.017
	(0.833)	(0.684)	(0.595)	(0.148)	(0.130)
Score Baseline Natural Science Test Conafe	8.105	8.061	8.105	-0.029	-0.006
	(0.715)	(0.561)	(0.608)	(0.122)	(0.125)
Score Baseline Social Sciences	7.872	7.894	7.965	0.057	0.085
	(0.723)	(0.644)	(0.624)	(0.129)	(0.130)
Repeater	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Panel B: Household Characteristics					
Indigenous Language	0.325	0.372	0.526	0.036	0.189
	(0.470)	(0.486)	(0.503)	(0.108)	(0.105)
Read	0.786	0.734	0.844	-0.043	0.058
	(0.412)	(0.444)	(0.365)	(0.065)	(0.059)
Less than Primary	0.624	0.479	0.462	-0.145	-0.166
	(0.486)	(0.502)	(0.502)	(0.084)	(0.070)
Upper Sec. or Higher	0.017	0.032	0.026	0.013	0.006
	(0.130)	(0.177)	(0.159)	(0.025)	(0.017)
Oportunidades	0.855	0.798	0.833	-0.065	-0.033
	(0.354)	(0.404)	(0.375)	(0.056)	(0.053)
Refrigerator	0.479	0.462	0.410	0.015	-0.046
	(0.502)	(0.501)	(0.495)	(0.078)	(0.089)
Television	0.795	0.806	0.718	0.021	-0.073
	(0.406)	(0.397)	(0.453)	(0.067)	(0.081)
Car	0.128	0.117	0.064	-0.015	-0.067
	(0.336)	(0.323)	(0.247)	(0.063)	(0.049)
Phone	0.256	0.298	0.244	0.057	-0.000
	(0.439)	(0.460)	(0.432)	(0.068)	(0.072)

Notes: School characteristics are based on the 2012 ENLACE and the 2013 Formato 911. Information on the characteristics of the community instructors draws on the 2016 follow-up survey, except for the number of training weeks that are reported in 2014 CONAFE administrative data. Student level information is based on the 2016 follow-up survey and the 2014 CONAFE administrative information on student diagnostics. Household level characteristics are obtained from the 2016 follow-up survey. The differences reported in the last two columns of the table are based on OLS regressions that control for stratification dummies. Standard errors for community instructors, student and household characteristics are clustered at school level.

Table E.9: Covariate Balance for Sub-Sample Above 7

Variable	(1) Control Mean (SE)	(2) API Standard Mean (SE)	(3) API Plus Mean (SE)	(4) Difference (2)-(1)	(3)-(1)
Panel A: Student Characteristics					
Baseline Age (Months)	105.049	104.536	105.834	-0.715	0.594
	(15.820)	(17.215)	(13.978)	(1.456)	(1.403)
Male	0.556	0.488	0.580	-0.066	0.029
	(0.498)	(0.501)	(0.495)	(0.041)	(0.048)
Scholarship	0.752	0.730	0.740	-0.022	-0.010
	(0.433)	(0.445)	(0.440)	(0.047)	(0.045)
Score Baseline Spanish. Test Conafe	7.275	7.463	7.309	0.193	0.026
	(0.839)	(0.848)	(0.710)	(0.119)	(0.100)
Score Baseline Math Test Conafe	7.397	7.535	7.349	0.146	-0.053
	(0.857)	(0.841)	(0.777)	(0.116)	(0.104)
Score Baseline Natural Science Test Conafe	7.378	7.553	7.357	0.180	-0.026
	(0.821)	(0.804)	(0.663)	(0.116)	(0.099)
Score Baseline Social Sciences	7.242	7.447	7.230	0.208	-0.021
	(0.843)	(0.800)	(0.690)	(0.118)	(0.105)
Repeater	0.033	0.005	0.017	-0.030	-0.019
	(0.178)	(0.069)	(0.128)	(0.015)	(0.018)
Panel B: Household Characteristics					
Indigenous Language	0.340	0.365	0.470	0.041	0.122
	(0.474)	(0.483)	(0.500)	(0.062)	(0.079)
Read	0.682	0.654	0.683	-0.031	0.002
	(0.466)	(0.477)	(0.466)	(0.050)	(0.052)
Less than Primary	0.621	0.649	0.641	0.025	0.014
	(0.486)	(0.478)	(0.481)	(0.049)	(0.049)
Upper Sec. or Higher	0.013	0.009	0.011	-0.004	-0.002
	(0.114)	(0.097)	(0.105)	(0.009)	(0.010)
Oportunidades	0.814	0.829	0.840	0.017	0.030
	(0.390)	(0.377)	(0.368)	(0.036)	(0.033)
Refrigerator	0.374	0.357	0.361	-0.015	-0.003
	(0.485)	(0.480)	(0.482)	(0.051)	(0.060)
Television	0.654	0.725	0.619	0.070	-0.031
	(0.477)	(0.448)	(0.487)	(0.052)	(0.058)
Car	0.069	0.066	0.061	-0.001	-0.006
	(0.253)	(0.249)	(0.240)	(0.023)	(0.024)
Phone	0.205	0.213	0.193	0.009	-0.003
	(0.404)	(0.411)	(0.396)	(0.042)	(0.041)

Notes: School characteristics are based on the 2012 ENLACE and the 2013 Formato 911. Information on the characteristics of the community instructors draws on the 2016 follow-up survey, except for the number of training weeks that are reported in 2014 CONAFE administrative data. Student level information is based on the 2016 follow-up survey and the 2014 CONAFE administrative information on student diagnostics. Household level characteristics are obtained from the 2016 follow-up survey. The differences reported in the last two columns of the table are based on OLS regressions that control for stratification dummies. Standard errors for community instructors, student and household characteristics are clustered at school level.

Table E.10: Total API Exposure at Scale and Original Treatment Assignment

Panel A: First Stage (Outcome: Transition to Secondary School Fall 2017)	
Original API Plus	1.96
	(0.10)
F-stat (Excluded Instrument)	414.94
Observations	625
Panel B: First Stage (Outcome: School is Open Fall 2018)	
Original API Plus	1.99
	(0.13)
F-stat (Excluded Instrument)	235.17
Observations	224

Notes: This table shows the first-stage estimates for the model in (2)-(3). The omitted category refers to the control group. In both panel A and panel B, the dependent variable in the first-stage regression represents the total years of API exposure. panel A shows the first-stage estimates when the second-stage repression outcome is transition to secondary school in the fall of 2017, while panel B shows the first-stage estimates when the second-stage repression outcome is whether schools are open in the fall of 2018.