The short-term effects of the Mobile Pedagogical Tutors: Evidence from a Randomized Control Trial in rural Mexico*

Francesco Agostinelli
University of Pennsylvania

Ciro Avitabile
University of Surrey
World Bank

Matteo Bobba
Toulouse School of Economics

Alonso Sanchez
World Bank

October 2018

Preliminary and Incomplete: Do not cite without permission

Abstract

In very remote and deprived areas students are hard to teach and teachers are difficult to train. We use data from a large Randomized Control Trial to evaluate the impact of Mobile Pedagogical Tutors on the education outcomes of students in 3rd to 6th grade in a community-based schooling model in rural Mexico, where the regular instructors are low educated non-professionals from the community. The mobile tutors devote their time providing a) pedagogical support to the instructors, b) one to one tutoring to the students lagging behind and 3) home visits to parents. Two years after the program inception, we find large and statistically significant effects on student assessments, parent reported measure of socio-emotional well-being and the probability of transition from primary to lower secondary. The effects are especially large when we augment the basic intervention with a strengthened training for the mobile tutors. When we study the channels through which the intervention affects student outcomes we find evidence that the mobile tutors led to an improvement of teachers' practices as well as parental investments. We do not find evidence that the intervention increased student effort.

*We would like to thank the Strategic Impact Evaluation Fund at the World Bank for the generous funding.

1 Introduction

Despite the dramatic progress towards the 2015 Education For All (EFA) goals for universal access to education (UNESCO 2015), there are still about 120 million children out of school, many of whom reside in rural areas. Even when they can stay in school, children in rural areas learn much less than their peers in urban areas. Students in rural areas miss many of the inputs that are necessary to achieve adequate levels of attainment and learning, both at home and in school. Demand based explanations can be ascribed, among others, to liquidity constraints, insufficient school readiness, lack of information, and social norms among parents. On the supply side, issues related to the general quality of infrastructures, and teachers in particular, are among the most widely cited barriers. Recruiting, preparing, retaining and monitoring teachers who have to work in remote areas is a daunting task for governments in many developing countries. Teacher quality issues in rural areas can be potentially exacerbated by limited accountability. Parents often do not have the tools to identify their children's learning needs and/or do not appreciate the extent to which the school system can effectively address those.

In this paper, we study how an intervention that addresses both demand and supply related factors affects cognitive and socio-emotional skills, as well as the educational attainments, of students in very remote and marginalized areas of Mexico. We use a large-scale field experiment to study whether and how a multifaceted pedagogical approach can improve student outcomes. The intervention combines one to one tutoring to worst performing students, pedagogical support to low-skilled teachers and home visits to spur parental involvement. The context is a community school model in rural Chiapas, the Mexican state where 77 percent of the population is classified as poor (CONEVAL 2016), the highest in the country.

Started in 1971, the *Consejo Nacional para el Fomento de la Educacion* (henceforth CONAFE) is a semi-autonomous government agency that is responsible for providing educational services in rural communities with fewer than 500 inhabitants. CONAFE's primary and secondary schools typically have a single multigrade classroom, with an average of 10-15 students who are taught by an instructor. The instructors, usually secondary school graduates, have no formal teaching qualification and they receive very basic training. Due to geographic remoteness, instructors have to move into the community, in exchange of a monthly allowance, and the promise of scholarship to continue their studies at tertiary education level. Not surprisingly, the results of CONAFE students are dramatically worse than those in the regular system and the gap has been widening over time. In 2007, 5 percent of the CONAFE students scored "Good or Excellent" in the national standardized test

ENLACE, as opposed to 22 percent in the Ministry of Education (SEP from the Spanish acronym). In 2013, 13 percent of CONAFE students scored Good or Excellent, vis a vis 42 percent in SEP schools.

In 2009, CONAFE started a mobile tutoring program aimed at enhancing the quality of education in these remote and disadvantaged schools. The tutors, known as API after the Spanish acronym, are recent university graduates who are hired on a temporary basis to provide educational support services. Each API is assigned to two schools, and they are expected to spend about two weeks per month in each school over a period of two years. According to the program guidelines, the tutors are supposed to 1) provide one-to-one tutoring to the worst performing students, 2) encourage parental involvement through home visits and 3) provide pedagogical support to the local teachers.

In 2014, as part of a World Bank project, we started a collaboration with CONAFE in order to evaluate and potentially improve the effectiveness of the API model. As part of this collaboration, we designed and evaluated an alternative training modality for the APIs with two main innovations. First, we doubled the initial training from one to two weeks, with the second week dedicated to hands one tutoring practices in reading and math. Second, we introduced bi-monthly meetings where the APIs could discuss with each other the challenges they faced and share ideas for solutions. We evaluated both modalities through a stratified Randomized Control Trial in 230 schools that had never received the intervention: 60 schools were assigned to the modality with strengthened training (henceforth API Plus), 70 schools assigned to the standard one (API Standard) and 100 schools were assigned to the control group, with no mobile tutor.

In schools where the mobile tutors received extra training and had bi-monthly meetings, the progression rate from primary to lower secondary education rose by 14 percentage points, which corresponds to an increase of 25 percent with respect to the average progression rate among students in control schools. Student learning in reading and math, as measured by the Early Grade Reading Assessment and the Early Grade Math Assessment respectively, increased by roughly 0.23 and 0.15 standard deviations (sd) when compared with the control group. Socio-emotional skills, as measured by parent reported measures, improved by 0.20sd. Students in schools that received the standard tutoring program showed much smaller improvements, although the difference between the two treatments is not always significant at conventional levels. Differences between the two treatment types are particularly large when we focus on the math and reading outcomes of students who were eligible for one to one tutoring, thus suggesting that additional teacher training is particularly beneficial for the most disadvantaged students. The improvements were primarily driven by changes in teachers' pedagogical practices and parental investments. We find no evidence that students changed their study effort.

These findings are important for two reasons. First, with a few notable exceptions, the interventions that have been found effective at improving learning outcomes in disadvantaged areas are potentially too expensive to be implemented on a large scale.[1] We provide evidence that a cost-effective multi-faceted intervention that addresses both demand and supply-related factors, can contribute to reduce the gap in educational attainments, cognitive and socio-emotional skills among children in highly marginalized areas. With a monthly salary of MXN $ 5,000, the APIs on average receive almost 40 percent less than teachers in the regular system while reaching a similar number of students.

Second, while there is growing evidence on the importance of teacher quality on student test scores (Steven G. Rivkin, 2005; Kane and Staiger, 2008) and longer run outcomes (Chetty et al., 2014b; Jackson, 2012), there is still little evidence on effective policies that improve teacher quality (Jackson et al., 2014). The literature is basically silent on how to improve teacher quality in rural and highly marginalized contexts. Low quality is the result of different factors that interact with each other. Attracting and retaining teachers in rural areas is extremely difficult, especially when teachers are not familiar with the social and cultural background. Teacher monitoring and supervision in rural areas often fail because of logistic issues. We have shown that a combined approach that provides pedagogical support to both students and teachers can be effective in improving the quality of teaching practices.

The paper is organized as follows. After reviewing the related literature, in Section 3 we provide some background on the CONAFE community model and the API tutoring model, as well as on the Chiapas pilot. Section 4 describes the data and the empirical specification. Section 5 presents the main results, as well as some of the heterogeneous effects. Section 6 discusses the potential channels through which the intervention can potentially affect student outcomes. Section 7 concludes.

2 Related Literature

This project builds on three previous classes of studies: the literature that assess the impact of education policies in poorly performing schools, the child development literature and the school-value added literature.

Recent attempts aimed at improving students' academic achievement in poorly performing schools revolve around the support of parents (e.g. Attanasio et al. (2014); Fernald et al. (2017)), teachers' training (e.g. Ozler et al. (2016); Yoshikawa et al. (2015)), and remedial

[1] As part of these few exceptions, see the computer assisted learning program Mindspark in India and a growth mindset program that changed beliefs and mental models of low income students in Peru.

education programs whereby underperforming kids work outside of the regular classroom with an extra-teacher, or tutor (e.g. Banerjee et al. (2007)). To the extent that the above inputs are complementary in the production function for academic achievement, we expect the combined effect of these policies to be larger than the sum of each individual component.

The child development literature focuses on the technology of skill formation as the object of interest, with the goal to understand optimal age of policy interventions in disadvantage children and its effects (see Cunha and Heckman, 2008; Cunha et al., 2010; Del Boca et al., 2014; Agostinelli and Wiswall, 2016). Todd and Wolpin (2003) and Todd and Wolpin (2007) considered a dynamic skill production function which includes past endowments as well as the sequence of family and school inputs supplied during the developmental period. Todd and Wolpin (2007) estimated this educational production function for children combining data from the C-NLSY79 (children of the respondent mothers of the original NLSY79 sample) together with three different school data sources: the Common Core Data, the School and Staffing Survey and the American Federation of Teachers. This data choice was needed to overcome the lack of information on child's school and classroom in C-NLSY79.

A prominent method to evaluate school and teacher effects on children achievement simply focuses on the impacts on student's test scores (e.g. Rivkin et al., 2005; Rockoff, 2004; Aaronson et al., 2007; Chetty et al., 2014a,b). Chetty et al. (2014b) estimate the effects of teacher quality on children outcomes in adulthood using teaching value added (VA) measures. They find substantial effects on future children income: exposing a classroom to an average teacher respect to a bottom 5 percent teacher in terms of VA would increase the present value of students lifetime income by approximately $250,000 per classroom.

A second contribution of this paper is to exploit the experimental variation generated from the intervention within a framework of multidimensional skill formation and endogenous parental and children's self investments.Jackson (2012) consider teacher effects on both test scores and non-test-score outcomes. Using North Carolina Administrative data, Jackson (2012) shows that teachers have effects on different outcomes like absences, suspensions, grades, and on-time grade progression. The author finds effects persist and they predict longer-run effects on high-school completion and college attendance. Flèche (2017) finds large effects on student's non-cognitive skills of teacher quality using UK birth cohort study. Additionally, Flèche (2017) finds teacher value-added on achievement test scores are poor predictors of teacher effects on non-cognitive skills. Finally, Fu and Mehta (2016) analyze how parental investments and classroom quality (through ability tracking regimes) interact in the development of children's skills.

3 Context and Research Design

3.1 The CONAFE community model

CONAFE is the Ministry of Education's decentralized agency responsible for providing educational services in rural communities with fewer than 500 inhabitants. In those communities, CONAFE offers the whole set of education services from early childhood development (age 6 months to 3 years and 11) until the end of lower secondary (9th grade). In year 2014, CONAFE schools accounted for roughly 10 percent of the about 99 thousand primary schools and 7 percent of the 38 thousands lower secondary schools in Mexico. In each level, CONAFE schools typically have a single multigrade classroom.[2]

The class has an average of 10-15 students, with students from the same grade usually grouped in circles, and is taught by a community instructor, called *Lider para la Educacion Comunitaria* (or LEC). Instructors must be 15-29 years old and have finished at least lower secondary school. They must be willing to move into a rural community to teach for at least a school year. They receive a stipend of MXN $1,427 per month. After one year of service, instructors receive a scholarship of MXN $982 per month for up to 30 months. This scholarship is conditioned on enrolling in a higher education institution. 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).

Dropout among community instructors is high. In the school year 2012-2013, 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 report having slept in the school, and 62 percent said they spent money on food.

3.2 The API Model

In 2009, CONAFE launched the "Mobile Pedagogical Tutors" program - *Asesores Pedagogicos Itinerantes* (API) - as an attempt to improve the quality of service delivery in the most rural schools. The APIs are selected among recent university graduates, preferably from pedagogy, psychology, sociology and social services related degrees, and with previous

[2]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 the CONAFE school.

experience as community instructors.[3] Until the school year 2015-2016 the knowledge of an indigenous language would be among the preference criteria. The API are usually hired for a two year period and receive a MXN $ 6,000 monthly salary. Partly due to the fixed-term contract, they are not covered either by social security or healthcare assistance.[4] Both the monetary and non-monetary benefits are significantly less generous than those available for teachers under the regular system, who have an permanent contract with a monthly salary starting from MXN $ 8,000 and are covered both by healthcare and social security provisions.

The APIs alternate their time on two-week intervals in two nearby school communities (School A and School B) during the academic year. Until the school year 2014-2015, A-type schools were selected according to the following criteria: a) at least 30 percent of the students classified as Insufficient in the Nationwide Standardized test ENLACE; b) 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. B-type schools were chosen exclusively on the basis of proximity to schools of type A. Similarly to the local instructors, during their service the APIs live in the two communities, which were supposed to take care of their basic needs (lodging and meals).

During their time spent in each of the communities, the APIs carry out three main activities with a predetermined time allocation. Specifically, they i) provide one to one tutoring to the worst-performing students in remedial sessions (60 percent of their time), ii) provide pedagogical support to teachers (15 percent), and iii) visit parents at their homes to provide them with information on their children's progress in school, and promote their participation in the school activities (25 percent). Each API is assigned to a maximum of six students for personalized tutoring, which in principle should take place outside of the regular school hours. During the regular school hours, the API is supposed to observe and take notes about the teaching practices of the local instructor, help him with the students who have learning difficulties and work outside of the classroom with those students who cannot attend one to one tutoring in the afternoon. Student eligibility for the one to one tutoring is determined by a joint assessment of the instructor and the API and it 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.[5]

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

[4]Starting from 2003, a revision of the General Health Law created the Popular Health insurance, designed for all those who are not insured otherwise.

[5]The evaluation covers the material that students should have mastered in the previous grade. It covers 7 subjects and the grades vary in the range between 5 and 10.

Preference is given to students in the 3rd to 6th grade. Once the eligible students have been identified, the API administers an additional exam, the so called *examen de colocación*, in order to establish the effective grade to which the student's knowledge corresponds. With this information in hand, the API prepares a personalized plan. Throughout the school cycle, the API provides the students with constant feedback and constantly monitors the progress of these students through a personalized evaluation form, the so called *Cuadernillo para el Seguimiento del Alumno.*

The objectives of the intervention go beyond the improvement of specific academic skills. Among others, the multifaceted approach aims at enhancing socio-emotional skills, such as self-esteem and self-regulation. Besides working on behavioral issues directly with the children, the APIs are supposed to address them also with parents as part of the home visits.

3.3 The Chiapas Pilot

In close collaboration with CONAFE in the year 2014, we developed and unrolled a pilot intervention in order to evaluate possible alternatives for strengthening the API model. Some of the insights for the pilot were drawn from a survey collected in 2011 in 40 communities in four states as part of a World Bank project. The main insights from the survey were the following: a) about two thirds of the tutors did not speak any indigenous language, although a large group of students reported an indigenous language as the only language spoken; b) only one third of the parents reported that the API had spent one hour or less doing home visits; c) one third of the APIs reported that the training module was not helpful in addressing the pedagogical challenges they were facing in the communities.

Based on the survey findings and conversations with former APIs, we designed two variants of the API model. The API Standard modality that would largely follow the standard model described in section 2.2, with two exceptions. First, the ability to speak the main indigenous language in the community would become the most important criterion for the assignment of APIs across the eligible communities. Second, the API supervisors would receive a salary increase in exchange of a mandatory increase in the frequency of their visits in the communities. The API Plus modality would include all the features of the API Standard, plus a significant change in the training module – i.e. two weeks rather than one week of training, with the second week focused on hands-on strategies to teach basic reading and math skills. In addition, the API Plus would receive four additional training sessions during the school year. These three day sessions (18 hours per session) allow the tutors to share experiences and design strategies jointly.

The pilot intervention was conducted in 230 schools of type A in rural Chiapas, that were randomly selected out of 260 schools that had never received the API intervention before.

The evaluation was conducted using a Stratified Randomized Control Trial, with the strata represented by the deciles of the 2012 school average ENLACE score and a randomization scheme performed separately within each ENLACE decile. As a result, 60 schools were assigned to the API Plus, 70 were assigned to the API Standard and 100 to the control group with no API intervention. The intervention was rolled out in August 2014 and the first data collection took place in Spring 2016. Throughout the whole implementation of the pilot program, we rely on administrative data collected by CONAFE in order to monitor the pilot's implementation. Six schools closed down their operations during the evaluation period: two of them due to the high level of political instability and the resulting social conflict in the communities while the remaining four because they were left with less than five students, the minimum required by CONAFE to leave a school open.

4 Data

4.1 Baseline Data

At the baseline no data collection was conducted. However, we rely on different administrative information in order to measure school and locality characteristics. From 2007 to 2013, ENLACE was administered to all students in the 3rd to 9th and 12th grades. The test is voluntary and has no effect on graduation or a student's GPA. The exam is administered in the students' schools by outside proctors. The score is normalized to have a mean of 500 and a standard deviation of 100. Nationwide, about 90 percent of the primary school students, although the average is significantly lower in CONAFE schools (about 60 percent). Previous work has found a strong correlation between the results in the primary school standardized tests and future academic outcomes, both in terms of attainment and learning. For this reason, we use the school average in the 2013 ENLACE test as a measure of school quality.

Information about school inputs (school and class size, student characteristics and teachers' credentials) was drawn from the census of schools carried out by the Secretariat of Public Education (Formato 911) twice a year. Using the census locality code, we retrieve information from the 2010 population census about the characteristics of the localities where the schools are located and from the National Commission for the Evaluation of Social Policy about the localities' poverty rate. In column 1 in Table 1 we report the average characteristics for the schools in the control group, while columns 2 and 3 report the p-values for the difference between the API Standard and the API Plus vis a vis the control group, as computed by a standard OLS regression that controls for strata fixed effects. and localities are reported in Table 1. Only for the probability of being a Primaria Indigena, out of 16 characteristics, we find a difference that is statistically significant at conventional level. The average test score for Spanish and math are about 0.9 and 0.7 standard deviations below the national average, but there are not statistically different in the three evaluation groups. Communities are on average small, 100 inhabitants, and are of difficult access, with 16 percent that have no road and 64 percent with no paved road. In 11 percent of the communities there are political conflicts (primarily due to the presence of Zapatista rebels).

4.2 Followup Data

In Spring 2016, a survey was conducted to measure the short term impacts of the API intervention on a variety of final outcomes and mediating inputs that might have been triggered by the intervention. The student module collected information among 1,930 children in 3rd to 6th grade. In order to measure the impact on cognitive skills, both the Early Grade Reading Assessment (EGRA) and the Early Grade Math Assessment (EGMA) were applied in order to measure the reading and basic math ability respectively. The EGRA is an individually administered oral student assessment that has been conducted in more than 40 countries and in a variety of languages. Typically administered in grade 2nd or 3rd, the interview includes timed assessments of letter naming, letter sound knowledge, phonemic awareness, pseudo-words and familiar words, oral reading fluency and untimed segments, including reading comprehension, listening comprehension and dictation. The EGMA is an individually administered oral assessment of foundational mathematic skills. The assessment includes counting, number identification, quantity discrimination measures, missing number measures, word problems, addition and subtraction, shape recognition and pattern extension, and it is usually applied to students in grades 1st to 3rd. Although in theory both EGRA and EGMA might be potentially too easy for children in higher grades, the comparison between 3rd graders in our control group and an urban sample suggests that the CONAFE children lag dramatically behind those in urban areas: more than 45 percent of the CONAFE students are more than 2 standard deviations below the urban average.

The module also collected information about student efforts, perceptions about parents' and community instructors' involvement and - in treatment communities - APIs' attention.

A household module collected information about 1,050 households, with an average of about 5 households per community. Due to the geographic dispersion, surveying the universe of households was outside the evaluation's budget and a random sample of households was selected within a 5km ray from the school. Besides standard information on household socioeconomic status, the module collected information on parents' expectations and investment into children's education including measures of homework supervision, interactions with teachers, time spent on school renovation, and number of books at home. In order to measure the impact of the intervention on socioemotional skills, we collected a caregiver reported measures of behavioral problems - a 32 items behavioral problem index (BPI) similar to the one applied in the NLSY79, that measures antisocial behavior, anxiety/depression, headstrongness, hyperactivity and peer conflicts.

Since APIs were not located in the community on a continuous basis, the survey firm interviewed them during the end of year evaluation session, that all APIs were supposed to attend. 107 APIs out of the 124 attended the sessions and were interviewed. Their characteristics are reported in Table 3, separately for those in the Standard and the Plus group. As a result of the randomization, API in the Standard group are not statistically different from those in the Plus group in terms of predetermined characteristics, such as gender, level of education and previous experience. There are also not statistically significant differences in terms of time spent in the community where the school A is located, and the number of students that receive one to one tutoring. Consistent with the differential intensity of the training in the two groups, the tutors in the Plus group report having attended more training (3.8 vs 2 weeks). The APIs in the Plus group are more likely helping the LECs in preparing the class and study cases, and less likely to prepare material - e.g. cutting pictures, preparing banners. The last piece of evidence is somehow suggestive that the APIs in the Plus group spend more time in activities that are more relevant for the quality of the pedagogical contents. When looking at activities they conduct with parents, the APIs in the Plus group are more likely to engage parents in activities related to learning and communication with their children.

5 Empirical Evidence

To estimate the causal impact of the two API modalities, we estimate the following equation:

$$Y_{ij} = \beta_0 + \beta_1 Standard_j + \beta_2 Plus_j + \delta T_j + \gamma' X_{ij} + u_{ij} \tag{1}$$

where Y_{ij} is the outcome of student i in school j recorded in the follow-up data. $Standard_j$ is an indicator dummy that takes the value one if school j is assigned to the API Standard group, 0 otherwise. $Plus_j$ is an indicator dummy that takes the value one if school j is assigned to the API Plus group, 0 otherwise. T_j are dummies to account for the 2012 ENLACE scores deciles that were used for the stratified randomization. β_1 and β_2 measures the Intention to Treat (ITT) effects of being assigned to the Standard and the Plus group. Let X_{ij} be a vector of baseline covariates, survey weeks and survey routes dummies. The last were made necessary by the fact that riots and particular climatic conditions took place only in specific weeks and specific places during the survey. In all the specifications, standard errors are clustered at school level to account for correlated shocks within schools, which represent the level at which the treatment is assigned.

Our main outcomes of interest are the scores in the EGRA, EGMA, the score in the behavioral program index rescaled in such a way that higher values are associated to fewer behavioral issues, and the dummy variable for transitioning from primary to lower secondary. We standardize all the scores using the mean and the standard deviation observed in the control group. When we study how the treatment effect varies along individual characteristics, the results are based on sample splits.

5.1 Average Effects

In Table 4 we report the results for the main outcomes of interest. In the odd numbered columns, we present the results that only control for strata dummies, survey week and survey routes fixed effects. In the even numbered columns we present the baseline specification, that also includes a dummy for sex, whether the child speaks indigenous language and the school average scores in math, Spanish and science as measured in 2013. The latter will be our baseline specification for the rest of the paper. Children in the schools that received the Standard API model saw an increase in their EGRA score by about 0.14 standard deviations (sd), as opposed to a 0.23 standard deviation improvement for those attending schools served by an API Plus. While larger, the coefficient for the API Plus is not statistically significant from the one for the API Standard. When we look at the performance in the test that measures math ability, we find no improvements for those in the API Standard group, while we see a moderately large and marginally significant effect for the API Plus.

Among others, the APIs were supposed to help students and parents identify the emotional issues that could affect their learning outcomes and deal with them. When we look at the impact on the variable that measures child socioemotional wellbeing, we find that the children in the API Plus saw an increase by 0.20 standard deviations, as opposed to a very small and statistically not significant effect for those in the API Standard. While the initial training did not have any specific content related to socioemotional issues, supervisors and coordinators encouraged API Plus to discuss with each other how to best deal with children's emotions during the bimonthly peer-to-peer sessions. This might explain the larger coefficient for the API Plus, although the difference between the two treatment groups is not statically significant at conventional levels.

Finally, we look at the probability of transitioning from primary to lower secondary for those who were enrolled in 6th grade during the school year 2015-2016. Upon graduating from a CONAFE school, children can enroll either into a CONAFE secondary school or a SEP school. Only 60 percent of the 6th graders in the control group transition from primary to lower secondary. Both children in the API Standard and in the API Plus model experience large increases in the probability of transitioning (10 percentage and 14 percentage point respectively), but only for the API Plus the effect is marginally significant at 10 percent. Although the effect sizes are larger for the API Plus than the API Standard for all the outcomes, we can never reject the null hypothesis of no differential effect.

In Table 5 we report the results for the EGRA and EGMA subdomains. When we look at the reading assessment (Panel A), we find that the intervention increases student proficiency in subdomains where they were already doing relatively well (familiar word reading, reading comprehension, reading correctly and dictation). There are no improvements in those subdomains where students where lagging most behind (initial sound and letter sound). This result is potentially consistent with two hypothesis. Sound acquisition happens relatively early and once children pass a certain age, it is much harder. The other hypothesis is that sound related questions were too demanding for children whose first mother tongue is an indigenous language might struggle to capture Spanish alphabet pronunciation. Panel B in Table 5 displays the results for EGMA. The only domain for which we can reject the null hypothesis of no impact of the API intervention is the one related to number identification. Addition and subtraction are among the sub-domains where children do worst at the baseline. For neither of them we observe any significant impact.

Overall, the results presented in Table 5 show that the API intervention improved children's competencies in areas where they were already doing relatively well, but not in those where they were lagging most behind.

5.2 Heterogeneous Treatment Effects

All children in CONAFE schools, irrespective of whether they receive or not the API intervention, are administered a diagnostic assessment at the beginning of the school year. In schools that receive the API intervention, the mobile tutor in collaboration with the LEC uses the results in the diagnostic test, plus information on whether the student is a repeater or not, and the LEC assessment in order to identify students eligible for one to one tutoring.[6]

[6]According to the program guidelines, everything else being equal, preference is given to 5th and 6th grade students in order to favor their transition to lower secondary. However, in our sample we do not observe differences in the percentage of students who are eligible for the one to one tutoring across grade.

The API applies a second test only to this set of students. Based on the results of this second test, 6 or fewer students are made eligible for the one to one tutoring.

In order to study the effect of the intervention on students who received the one to one tutoring and those who did not, we proceed in two steps. First, using the estimated coefficients of a logit regression of the eligibility dummy for the one to one tutoring on a dummy for student gender, age, GPA in the previous year, the dummy for speaking the indigenous language, the scores in the diagnostic test for Spanish and math at the beginning of the school year 2015-2016 and grade dummies, we compute the predicted probability of being eligible for the one to one tutoring among students in the control group. In each control school we rank students in terms of predicted probability of being at risk, and we compare the five ones with the highest probability with those in the treatment schools, that are eligible according to the administrative records.[7]

In Panel A in Table 6 we present separately the results for children who received the one to one tutoring and those who did not. When we look at the effect on cognitive skills, as proxied by reading and math ability, we can reject that the program had no impact only for the students who were eligible for the one to one tutoring, but not for the others. The effects seem to be produced primarily by the API Plus model, that displays seem much larger effects than the API Standard. The opposite is true when we look at the socio-emotional wellbeing score and the probability of transitioning to secondary school, although for the last outcome the number of observations is so limited that the evidence has to considered as merely suggestive.

These results seem to suggest that providing the APIs with extra training - especially hands-on teaching methods in math and reading, is improving math and reading skills of children who are most in need, possibly through the effect of one to one sessions. As mentioned before, the initial training had no special emphasis on socio-emotional skills and any potential explanation for the differential effect of the API Plus model would be merely speculative.

In Panel B in Table 6, we study the differentiated effect by whether the child speaks an indigenous language or not. Not surprisingly, the averages for the control group suggest that children with an indigenous background do much worse in math and reading than the non-indigenous ones, with a difference that is about 0.4 standard deviations in both. The transition from primary to lower secondary is instead higher indigenous than non-indigenous, perhaps as a result of reduced labor opportunities for the former. Overall, our results seem to suggest that there are no differences in the treatments' effectiveness depending on the

[7]On average, about 5 students among those eligible for the one to one tutoring in the treatment schools were attending.

language.

Last we look at the effects by grade. As already mentioned before, CONAFE schools are multigrade, with children from the same grade sitting in a circle around one table. According to the CONAFE classification, children in grades 3 and 4 are part of the Level 2, while children in grade 5 and 6 are part of the Level 3. When we study the effects separately for children in Levels 2 and 3, we can reject the null hypothesis that the API had no significant effect only for children in Level 2, but not for children in Level 3. In this case, we find evidence that the effect of the API Plus treatment is much larger than the API Standard for all the outcomes of interest. One potential explanation, for the small effects among students in grades 5 and 6 might be that the EGRA and EGMA are too easy, generating a potential ceiling effect. However when look at the share of correct answers for control group children attending these grades, we find the average share of correct answers increases with grade; but there is large scope for improvement among 5th and 6th graders (Fig.). We are therefore more inclined to interpret these results as evidence of the presence of critical age windows for the acquisition of basic skills.

Overall, the results presented in this section, consistent with the well established evidence on Early Childhood interventions, show that an intensive and direct intervention is more effective in early grades and can potentially benefit the most disadvantaged students.

6 Channels

In this section we provide evidence on the channels through which the intervention might have improved student outcomes. There at least three inputs that might have changed as a result of the intervention: 1) teachers' effectiveness, 2) parental investment, 3) student effort. Due to to the multifaceted nature of the intervention, we are not be able to isolate the importance of each of these three mechanisms. The intervention might have changed not only the amount of the inputs that contribute to improved child outcomes, but also their productivity. Distinguishing between these two potential mechanisms goes beyond the objectives of this paper.

6.1 Teachers' Effectiveness

There are two ways through which the mobile tutor might contribute to improve the LEC's effectiveness in the classroom. First, there is the *direct* effect of the pedagogical support. This might concretize in different ways: class preparation, joint class delivery, and feedback provision after the class. There might also be an *indirect* effect induced by the one to one tutoring: as the worst off students catch up with the average student, teacher's effectiveness increases. In order to study the effect of the API intervention, we use data from a modified version of the Stallings "classroom snapshot" observation system. The instrument was adjusted to account for the multigrade nature of the CONAFE schools. We focus on three aspects of teacher's pedagogy: 1) the ability to adjust to students' learning speed, as proxied by a dummy variable; 2) an index that proxies for the amount dedicated to teaching activities, as obtained by a principal component analysis of the time devoted to six activities;[8] 3) an index that proxies for the class engagement, as obtained by a principal component analysis for a subgroup of four activities that have a direct link to reading and numeracy skills. We analyze the effect of the intervention on each of these three variables, plus an overall score - obtained through the principal component of the three above and an index that summarizes material availability.

The results are presented in Table 7. LECs who were exposed to an API, irrespective of whether belonging to the API Plus or the API Standard modes, displayed a better ability to adjust to the student learning speed (column 1), as reported by the external observer. When we consider the effect on the scores that captures the amount of time that the LEC devotes to activities that have a more direct impact on student learning, and not activities such as classroom administration, discipline and social interactions that did not involve the student, we find that the API Plus had a large (0.28 standard deviations) but not statistically significant impact, while the API Standard had a negative although small and not statistically significant impact. We do find remarkably similar effects when we look at LEC's ability to engage students in activities that are more conducive of better reading and math skills. When we consider the overall score, we do find a large (0.35sd) and statistically significant effect for the Plus model, and a zero effect for the Standard one.

6.2 Parental Investment

The time and the material investments that parents devote to their children vary considerably with the socioeconomic background. Only part of the socioeconomic gradient in parental investment depend on parents' available resources. Parents from low socioeconomic background are more likely to have downwards biased beliefs about the returns to child investment and beliefs are strongly correlated with actual investment choices (Attanasio et al., 2018; Cunha et al., 2013). Home visits have been found to be an effective strategy in improving cognitive and socio-emotional outcomes of preschool age children, primarily through increased parental investment (Attanasio et al., 2015). There is no such a evidence for school

[8] The activities include reading loud individually, reading in group, demonstration, questions and answer, memorization, and individual homework.

age children. By providing home visits, the API assesses both the household environment and informs the parents about the importance of investing in their children and how they can contribute to their development.

We use information from the household questionnaire to assess how the two different modalities affected parental aspirations In Table 8 we present the results for a set of parental outcomes. Parents at the baseline have very low aspirations about their children's future education attainments: only 9 percent expect them to complete high school or higher. Among parents who are exposed to the the visit of on API Plus, the share doubles, while there is no change among those who receive the visit of an API Standard. We next look at the effect on 5 outcomes that proxy for parental investment: a) how often they help children with homework, with 1 denoting the minimum and 5 the maximum; b) how often they supervise homework, with 1 denoting the minimum and 5 the maximum, c) the number of meetings with teachers, d) the number of books at home, e) the number of school improvement activities. The results are reported in columns (2) to (6) in Table 8. Although not always statistically significant, results in columns (2) to (5) show that being exposed either to an API Plus or API Standard has always a positive effect. Only for the number of school activities, we find a positive effect only for the API Plus model. In order to summarize the overall impact on parental behavior, we use principal component to build a summary measures. The results are reported in column (7) in Table 8. Among parents exposed to API Plus, the summary measure of parental investment increases by 0.14sd. The effect is negative, although small and not statistically significant, among parents who were exposed to an API Standard. The difference between the two treatment types is statistically significant (p-value=0.046).

The results in this section suggest that parents changed their aspirations and their behavior about their children's schooling as a result of the exposure to the mobile tutor. So far we have emphasized the role of the home visits as potential channel through which the intervention can affect parental investments. However, we can not completely rule out that parents in the API Plus group might respond to the improvement in teaching practices and perhaps their children's learning by investing more.

6.3 Student Effort

In section 6.1 we provided evidence that the API intervention increased student engagement in math and reading related activities. Increased engagement might be simply the result of improved teaching techniques, but it might also depend on students increasing their effort. As students perceive that the quality of their learning environment improves, they might have more incentives to increase their effort.

We do not collect self-reported measures of time devoted to homework. We focus on the three outcomes, as reported by the students: i) the probability of working outside school time, ii) the number of days missed and iii) and the number of delays. Results are reported in Table 9. In columns (1), (4) and (7) we report the results for the overall sample. In columns (2), (5) and (8) we report the results for those who were not eligible for the one to one tutoring, while in columns (3), (6) and (9) we report the results for those eligible for the one to one tutoring. A priori, it is possible that only children who were directly exposed to the one to one sessions, increased their effort as they are the ones who potentially saw larger improvements. A very large share of children in our sample report working at the baseline (about 70 percent), and we do observe no change either for those who receive API Standard or an API Plus.

On average, children miss 4 days of school every month, with the baseline being much higher for those who are eligible for the one to one tutoring than for those who are not (5 vs 3). A similar pattern is found for the delays in the last month. Overall, the results presented in Table 9 suggest no impact on any proxy for student effort.

Taken together, the evidence presented in this section, while mostly speculative, would be consistent with the hypothesis that in an extremely poor and marginalized population, such as the one that is the object of our study, employment and attendance decisions depend primarily on household liquidity constraints and pre-existing readiness gaps, and less on the current quality of the school environment.

7 Conclusion

We provide evidence on the impact of a scalable mobile tutor program implemented in a highly marginalized context in rural Mexico. Mobile tutors spend two weeks per month in a school with very low achievements and the other two weeks in a geographically close community, and they are supposed to divide their time among three activities: one to one tutoring to students who lag behind, pedagogical support to community teachers, and home visits.

Using data from a survey collected two years after the program's inception and administrative sources, we find sizable effects on student attainments and cognitive outcomes. We also find a significant effect on a parent reported measure of emotional well being. The mobile tutor program is particularly effective when augmented with improved training, consisting of an extra week of training in the beginning of the year and peer to peer sessions on bimonthly basis.

Our results show that a low cost multifaceted intervention that addresses both demand

and supply related constraints can contribute to partly reduce the gap between children in urban and rural communities. There has been an increased attention towards program scalability. Both interventions tested during the pilot were designed according to CONAFE's available resources. As a result, a universal roll-out of the improved training model for the mobile tutor was started in Fall 2017.

References

Daniel Aaronson, Lisa Barrow, and William Sander. Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics*, 25(1):95–135, 2007.

Francesco Agostinelli and Matthew Wiswall. Estimating the technology of children's skill formation. *NBER Working Paper*, (22442), 2016.

Orazio Attanasio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. Estimating the production function for human capital: Results from a randomized control trial in colombia. *NBER Working Paper*, (21740), 2015.

Orazio Attanasio, Teodora Boneva, and Christopher Rauh. Parental Beliefs about Returns to Different Types of Investments in School Children. Working Papers 2018-032, Human Capital and Economic Opportunity Working Group, May 2018. URL https://ideas.repec.org/p/hka/wpaper/2018-032.html.

Orazio P Attanasio, 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 child development program in colombia: cluster randomized controlled trial. *BMJ*, 349, 2014. doi: 10.1136/bmj.g5785. URL https://www.bmj.com/content/349/bmj.g5785.

Rosangela Bando and Claudia Uribe. Experimental evidence on credit constraints. Working Paper 670, Inter-American Development Bank, February 2016.

Abhijit V. Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden. Remedying education: Evidence from two randomized experiments in india*. *The Quarterly Journal of Economics*, 122(3):1235–1264, 2007. doi: 10.1162/qjec.122.3.1235. URL http://dx.doi.org/10.1162/qjec.122.3.1235.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2014a.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–79, 2014b.

Flavio Cunha and James J. Heckman. Formulating, identifying, and estimating the technology for the formation of skills. *Journal of Human Resources*, 43(4):738–782, 2008.

Flavio Cunha, James J. Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, 2010.

Flvio Cunha, Irma Elo, and Jennifer Culhane. Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation. NBER Working Papers 19144, National Bureau of Economic Research, Inc, June 2013. URL https://ideas.repec.org/p/nbr/nberwo/19144.html.

Daniela Del Boca, Christopher Flinn, and Matthew Wiswall. Household choices and child development. *Review of Economic Studies*, 81(1):137–185, 2014.

Lia C H Fernald, Rose M C Kagawa, Heather A Knauer, Lourdes Schnaas, Armando Garcia Guerra, and Lynnette M Neufeld. Promoting child development through group-based parent support within a cash transfer program: Experimental effects on children's outcomes. *Developmental psychology*, 53(2):222—236, February 2017. ISSN 0012-1649. doi: 10.1037/dev0000185. URL https://doi.org/10.1037/dev0000185.

Sarah Flèche. Teacher quality, test scores and non-cognitive skills. *CEP Discussion Papers*, 2017.

Chao Fu and Nirav Mehta. Ability tracking, school and parental effort, and student achievement: A structural model and estimation. *Working Paper*, 2016.

C. Kirabo Jackson. Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in north carolina. Working Paper 18624, National Bureau of Economic Research, December 2012.

C. Kirabo Jackson, Jonah E. Rockoff, and Douglas O. Staiger. Teacher effects and teacher-related policies. *Annual Review of Economics*, 6(1):801–825, 2014.

Thomas J Kane and Douglas O Staiger. Estimating teacher impacts on student achievement: An experimental evaluation. Working Paper 14607, National Bureau of Economic Research, December 2008.

Berk Ozler, 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. URL https://ideas.repec.org/p/wbk/wbrwps/7817.html.

Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005.

Jonah E. Rockoff. The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94(2):247–252, 2004.

John F. Kain Steven G. Rivkin, Eric A. Hanushek. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, March 2005.

Petra E. Todd and Kenneth I. Wolpin. On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113(485):F3–F33, 2003.

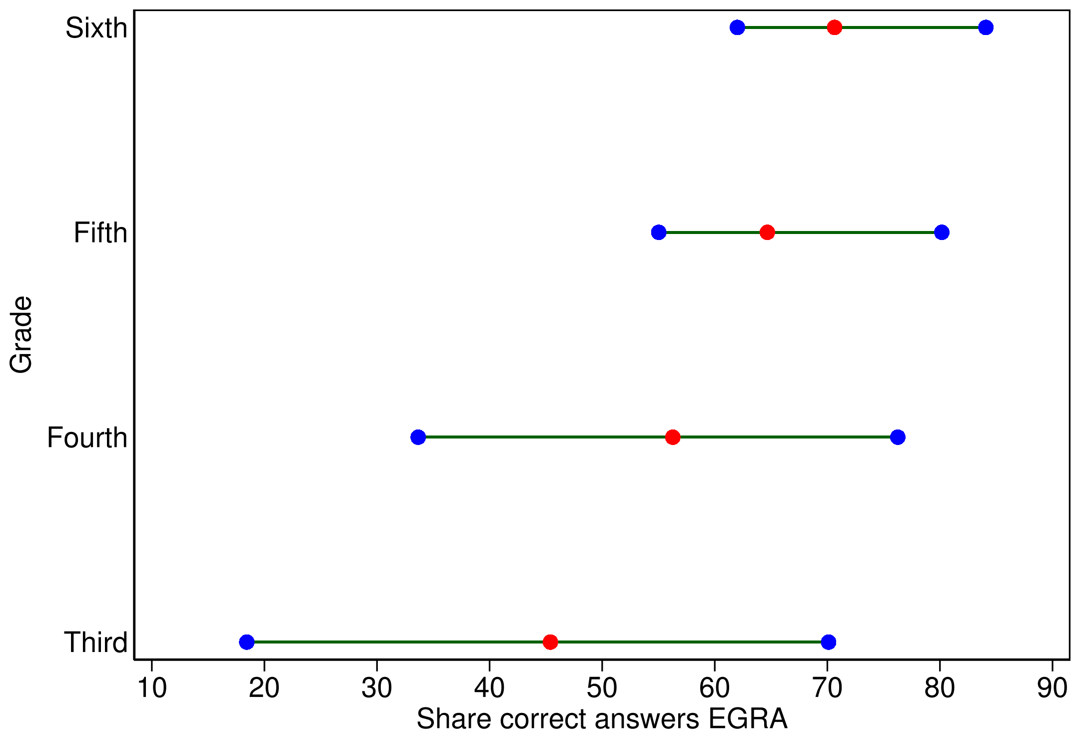
Petra E. Todd and Kenneth I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, 1(1):91–136, 2007.

H. Yoshikawa, Snow Leyva, D., E. C. E., Trevino, M. C. Barata, C. Weiland, and M. C. Arbour. Experimental impacts of a teacher professional development program in chile on preschool classroom quality and child outcomes. *Developmental psychology*, 51(3):309—322, March 2015.

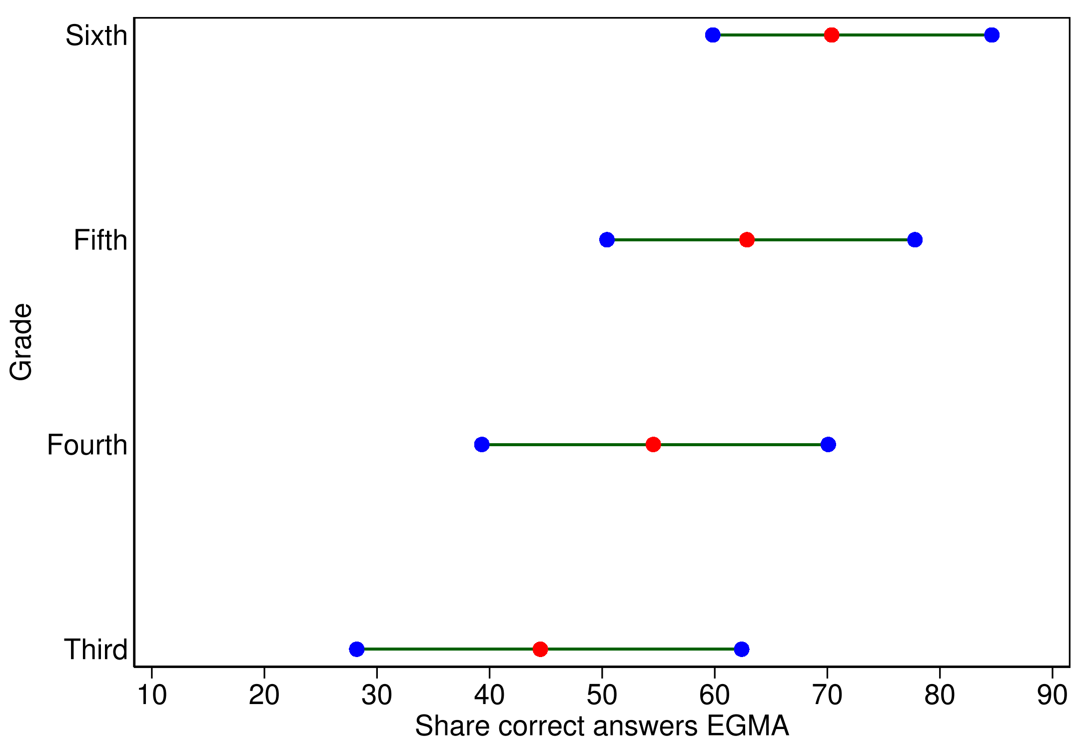
Table 1: Baseline Characteristics

	Control	Standard	Plus	Standard=Control p-value	Plus=Control p-value	F test	Obs
Rooms in Use	1	1	1				226
	0	0	0				
Rooms in the School	1	1	1				226
	0	0	0				
Teachers	1.22	1.309	1.207	0.176	0.856	1.493	226
	0.416	0.465	0.409				
Enrollment	15.19	15.441	14.379	0.789	0.359	1.296	226
	5.81	5.655	5.824				
ENLACE Spanish	428.433	432.326	430.573	0.763	0.826	296.43	226
	60.244	67.579	67.463				
ENLACE Math	452.381	455.82	451.627	0.983	0.691	64.498	226
	77.986	84.546	82.461				
ENLACE Science	438.232	441.259	442.856	0.843	0.464	49.81	226
	49.87	49.323	50.492				

Figure 1: EGRA and EGMA by Grade



The blue dots denote the bottom 20 and top 20 percentiles in the control group. The red dot denotes the control group mean.



The blue dots denote the bottom 20 and top 20 percentiles in the control group. The red dot denotes the control group mean.

Table 2: Followup Student and Household Characteristics

	Control	Standard	Plus	Standard=Control p-value	Plus=Control p-value	F test	Obs
				Student Characteristics			
Baseline Age Months	107.915	107.145	107.994	0.36	0.984	1.256	1922
	17.431	18.744	16.976				
Male	0.504	0.522	0.518	0.464	0.597	0.565	1930
	0.5	0.5	0.5				
Indigenous Language	0.313	0.329	0.453	0.753	0.035	4.377	1930
	0.464	0.47	0.498				
Scholarship	0.732	0.738	0.757	0.771	0.435	0.964	1930
	0.443	0.44	0.429				
Baseline Conafe Info	0.819	0.872	0.857	0.234	0.358	1.711	1930
	0.385	0.334	0.351				
CONAFE Spanish Diagn.	7.682	7.688	7.68	0.896	0.944	0.837	1725
	0.881	0.795	0.81				
CONAFE Math Diagn.	7.684	7.656	7.688	0.845	0.956	0.884	1725
	0.904	0.818	0.821				
CONAFE NatSciences Diagn.	7.77	7.747	7.733	0.859	0.607	0.793	1725
	0.869	0.775	0.744				
CONAFE SocialForm Diagn.	7.585	7.597	7.569	0.861	0.814	1.158	1725
	0.843	0.737	0.78				
CONAFE AverageScore	7.888	7.92	7.881	0.698	0.87	0.835	1725
	0.81	0.687	0.674				
				Parent Characteristics			
Indigenous Language	0.329	0.366	0.476	0.501	0.065	3.767	1053
	0.47	0.483	0.5				
Read	0.711	0.686	0.734	0.514	0.556	1.8	1050
	0.454	0.465	0.443				
Less than Primary	0.619	0.587	0.584	0.441	0.428	1.791	1053
	0.486	0.493	0.494				
Upper Secondary or Higher	0.015	0.016	0.019	0.999	0.687	1.688	1053
	0.122	0.124	0.135				
				Household Characteristics			
Oportunidades	0.81	0.807	0.829	0.974	0.579	0.572	1053
	0.393	0.395	0.377				
Refrigerator	0.397	0.388	0.373	0.834	0.701	1.455	1049
	0.49	0.488	0.485				
Television	0.69	0.738	0.651	0.293	0.455	1.056	1052
	0.463	0.44	0.478				
Car	0.084	0.081	0.063	0.898	0.382	1.445	1052
	0.278	0.273	0.244				
Sewage	0.25	0.253	0.32	0.941	0.177	0.459	1049
	0.433	0.435	0.467				
Phone	0.22	0.233	0.204	0.688	0.702	2.594	1050
	0.415	0.423	0.404				
Light	0.859	0.916	0.873	0.144	0.74	1.057	1052
	0.348	0.278	0.333				

Table 3: Followup API Characteristics

	(1) API Plus	(2) API	(3) p-value	(4) p-value Perm	(5) Obs
Age	28.4	28.386	0.709	0.713	107
	3.057	3.678			
Male	0.62	0.579	0.597	0.597	107
	0.49	0.498			
High Edu Complete	0.88	0.877	0.926	0.922	107
	0.328	0.331			
Training Weeks Current	3.86	2.105	0	0	107
	2.339	1.332			
Training Weeks Previous Cycle	3.14	2	0.057	0.056	95
	3.036	2.039			
Experience as Api	20.04	22.298	0.269	0.29	107
	8.755	10.997			
LEC or Tutor	0.48	0.25	0.014	0.016	106
	0.505	0.437			
Education Assistant	0.06	0.07	0.994	0.985	107
	0.24	0.258			
Days in Community A	14.26	13.298	0.436	0.444	107
	5.645	4.822			
Days Stay A	7.48	7.842	0.585	0.587	107
	6.238	6.253			
Students Lagging Behind	5.76	5.789	0.852	0.856	107
	1.847	1.521			
Time students lagging behind (min)	154.286	130.842	0.555	0.579	106
	189.217	167.662			
Help LEC with Planning (Y/N)	0.6	0.439	0.096	0.093	107
	0.495	0.501			
Help LEC with study cases (Y/N)	0.9	0.807	0.196	0.205	107
	0.303	0.398			
Help LEC pers sess (Y/N)	0.9	0.877	0.771	0.765	107
	0.303	0.331			
Help LEC prepare material (Y/N)	0	0.158	0.005	0.002	107
	0	0.368			
Time with Parents (min)	134.4	140	0.633	0.626	107
	85.336	97.761			
Topic Home Visit: Culture	0.78	0.789	0.814	0.813	107
	0.418	0.411			
Topic Home Visit: Communication	0.86	0.702	0.047	0.047	107
	0.351	0.462			
Topic Home Visit: Feelings	0.54	0.509	0.752	0.736	107
	0.503	0.504			
Topic Home Visit: Self	0.6	0.596	0.979	0.98	107
	0.495	0.495			
Topic Home Visit: Learning	0.74	0.596	0.098	0.101	107
	0.443	0.495			

Table 4: Main Impacts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	EGRA		EGMA		Socioemotional Score		Transition Primary to Lower Sec.	
API Stantdard	0.116	0.135**	0.008	0.025	0.067	0.080	0.090	0.103
	(0.074)	(0.065)	(0.071)	(0.065)	(0.086)	(0.088)	(0.073)	(0.073)
API Plus	0.197**	0.227***	0.121	0.145*	0.203**	0.204**	0.125	0.145*
	(0.083)	(0.078)	(0.084)	(0.081)	(0.088)	(0.087)	(0.076)	(0.075)
Strata F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
H0: Standard+Plus=0	0.049	0.010	0.304	0.174	0.072	0.065	0.216	0.124
H0: Standard=Plus	0.348	0.238	0.177	0.112	0.154	0.203	0.674	0.615
Mean Dep. Var.	0.004	0.004	0.004	0.004	-0.000	-0.000	0.602	0.602
SD Dep. Var.	0.995	0.995	0.995	0.995	1.000	1.000	0.491	0.491
Adjusted R-squared	0.128	0.176	0.089	0.125	0.066	0.071	0.008	0.037
Observations	1895	1895	1895	1895	1037	1037	385	385
Clusters	224	224	224	224	221	221	169	169

Table 5: Analysis of EGRA and EGMA by subdomain

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Share of Correct Answers EGRA by Sub-Domain									
	Letter name	Initial sound	Letter sound	Familiar word r.	Non-word reading	Reading comp.	Reading correctly	Listening comp.	Dictation	weighted % Average
API Stantdard	2.509	0.978	2.379	4.223*	2.866	4.631**	5.135**	-0.151	2.727	2.811*
	(1.707)	(2.369)	(1.536)	(2.145)	(1.917)	(2.319)	(2.289)	(2.195)	(1.908)	(1.507)
API Plus	3.795**	-2.134	0.919	6.315**	4.483**	7.325**	7.923***	3.319	7.419***	4.374**
	(1.897)	(2.631)	(1.451)	(2.448)	(2.209)	(2.929)	(2.429)	(2.463)	(1.959)	(1.783)
H0: Standard+Plus=0	0.097	0.516	0.303	0.023	0.091	0.027	0.003	0.277	0.001	0.033
H0: Standard=Plus	0.538	0.255	0.367	0.407	0.497	0.351	0.290	0.136	0.024	0.388
Mean Dep. Var.	60.422	28.322	29.279	71.104	58.861	65.816	73.513	56.478	61.229	56.114
SD Dep. Var.	25.711	32.183	22.212	34.544	31.322	39.129	36.057	32.321	28.810	23.273
Adjusted R-squared	0.243	0.101	0.088	0.192	0.202	0.232	0.219	0.278	0.191	0.260
Observations	1895	1895	1895	1895	1895	1895	1895	1895	1895	1895
Clusters	224	224	224	224	224	224	224	224	224	224
	Panel B: Share of Correct Answers EGMA by Sub-Domain									
	Number Identification	Number Discrimination	Missing Numbers	Addition	Subtraction	Problem solving	Shapes Recognition	weighted % Average		
API Stantdard	2.904*	-0.848	1.278	-0.279	0.500	-0.939	-0.220	0.342		
	(1.665)	(1.386)	(1.914)	(1.541)	(1.360)	(1.675)	(1.771)	(1.207)		
API Plus	4.585**	2.335*	2.565	1.195	-0.347	1.424	1.736	1.928		
	(2.012)	(1.404)	(2.256)	(1.740)	(1.311)	(2.097)	(2.080)	(1.475)		
H0: Standard+Plus=0	0.055	0.144	0.518	0.695	0.840	0.439	0.600	0.393		
H0: Standard=Plus	0.392	0.063	0.556	0.411	0.557	0.205	0.333	0.250		
Mean Dep. Var.	73.243	84.669	48.298	53.560	35.400	53.586	68.144	59.557		
SD Dep. Var.	28.476	21.275	30.295	26.182	23.487	30.082	24.339	19.771		
Adjusted R-squared	0.186	0.116	0.145	0.245	0.142	0.192	0.173	0.247		
Observations	1895	1895	1895	1895	1895	1895	1895	1895		
Clusters	224	224	224	224	224	224	224	224		

Table 6: Treatment Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	EGRA		EGMA		Socio Emot. Score		Transition to Lower Sec	
	Panel A: Heterogeneity by Targeting of the 1to1 tutoring							
	Not Targeted	Targeted	Not Targeted	Targeted	Not Targeted	Targeted	Not Targeted	Targeted
API Stantdard	0.153**	0.158*	0.093	0.056	0.154	-0.012	0.168*	-0.126
	(0.071)	(0.089)	(0.081)	(0.087)	(0.127)	(0.110)	(0.086)	(0.101)
API Plus	0.152*	0.330***	0.065	0.286***	0.290**	0.113	0.266***	0.024
	(0.086)	(0.101)	(0.102)	(0.101)	(0.120)	(0.120)	(0.096)	(0.104)
H0: Standard+Plus=0	0.057	0.005	0.503	0.013	0.053	0.546	0.016	0.272
H0: Standard=Plus	0.996	0.078	0.793	0.015	0.336	0.312	0.323	0.137
Mean Dep. Var.	0.128	-0.157	0.163	-0.198	-0.063	0.048	0.570	0.647
SD Dep. Var.	0.940	1.058	0.942	1.030	1.017	0.995	0.498	0.481
Adjusted R-squared	0.183	0.190	0.154	0.131	0.045	0.066	0.192	0.084
Observations	835	975	835	975	439	569	181	182
Clusters	179	218	179	218	166	213	113	117
	Panel B: Heterogeneity by Indigenous Language							
	Not Indig.	Indig.	Not Indig.	Indig.	Not Indig.	Indig.	Not Indig.	Indig.
API Standard	0.120**	0.169	0.017	0.017	0.153	0.019	0.164*	0.101
	(0.056)	(0.132)	(0.063)	(0.122)	(0.094)	(0.173)	(0.085)	(0.139)
API Plus	0.227***	0.254**	0.192**	0.050	0.175*	0.214	0.286***	0.007
	(0.081)	(0.126)	(0.094)	(0.124)	(0.105)	(0.146)	(0.096)	(0.122)
H0: Standard+Plus=0	0.010	0.129	0.111	0.913	0.135	0.292	0.009	0.744
H0: Standard=Plus	0.195	0.515	0.059	0.761	0.847	0.258	0.243	0.513
Mean Dep. Var.	0.141	-0.303	0.124	-0.262	0.004	-0.008	0.569	0.667
SD Dep. Var.	0.942	1.042	1.002	0.926	1.008	0.986	0.497	0.475
Adjusted R-squared	0.120	0.193	0.090	0.109	0.093	0.029	0.119	0.210
Observations	1246	649	1246	649	688	349	245	140
Clusters	179	125	179	125	174	107	122	70
	Panel C: Heterogeneity by Level (3rd and 4th grade vs 5th and 6th grade)							
	Grade 3 and 4	Grade 5 and 6	Grade 3 and 4	Grade 5 and 6	Grade 3 and 4	Grade 5 and 6		
API Standard	0.137*	0.149*	0.034	0.028	0.040	0.087		
	(0.073)	(0.081)	(0.074)	(0.084)	(0.098)	(0.145)		
API Plus	0.312***	0.119	0.224**	0.045	0.212**	0.158		
	(0.091)	(0.093)	(0.093)	(0.097)	(0.096)	(0.132)		
H0: Standard+Plus=0	0.003	0.162	0.040	0.884	0.069	0.471		
H0: Standard=Plus	0.057	0.740	0.028	0.866	0.087	0.676		
Mean Dep. Var.	0.000	0.008	0.002	0.007	0.055	-0.081		
SD Dep. Var.	1.001	0.990	1.001	0.990	1.024	0.961		
Adjusted R-squared	0.222	0.141	0.167	0.108	0.082	0.053		
Observations	1001	894	1001	894	617	420		
Clusters	217	213	217	213	212	188		

Table 7: Potential Channels: Teachers' Pedagogy

	(1) Keep Student Learning Speed (Y/N)	(2) Teaching Activities Score	(3) Student Engagement Score	(4) Overall Score
API Standard	0.076*	-0.048	-0.066	-0.075
	(0.043)	(0.180)	(0.179)	(0.185)
API Plus	0.096**	0.282	0.285	0.356**
	(0.041)	(0.175)	(0.174)	(0.180)
H0: Standard+Plus=0	0.040	0.181	0.157	0.073
H0: Standard=Plus	0.664	0.101	0.080	0.039
Mean Dep. Var.	0.069	0.079	0.079	0.071
SD Dep. Var.	0.255	0.271	0.271	0.258
Adjusted R-squared	0.068	0.171	0.189	0.172
Observations	199	203	203	194

Table 8: Potential Channels: Parental Investments

	(1) Would like child complete upper sec. or higher	(2) How often parent helps with homework [1-5]	(3) How often parent verifies homework [1-5]	(4) Meetings with teachers	(5) Books at home	(6) School Improvement Activities	(7) Parental Investment Factor
API Standard	0.006	0.188*	0.161**	0.352	3.098*	-0.146	-0.027
	(0.036)	(0.099)	(0.076)	(0.506)	(1.597)	(0.097)	(0.063)
API Plus	0.094**	0.216*	0.073	1.235**	1.950	0.157	0.144**
	(0.039)	(0.118)	(0.092)	(0.561)	(1.521)	(0.098)	(0.071)
H0: Standard+Plus=0	0.039	0.085	0.110	0.091	0.133	0.034	0.084
H0: Standard=Plus	0.034	0.821	0.377	0.172	0.504	0.009	0.041
Mean Dep. Var.	0.088	0.087	0.087	0.085	0.091	0.087	0.088
SD Dep. Var.	0.283	0.283	0.283	0.280	0.288	0.283	0.284
Adjusted R-squared	0.116	0.120	0.165	0.030	0.100	0.129	0.216
Observations	1008	1036	1034	965	865	1037	821
Clusters	221	221	221	219	214	221	212

Table 9: Potential Channels: Student Effort

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Works (Y/N)			Days Missed			Delays	
	Full Sample	Not Targeted	Targeted	Full Sample	Not Targeted	Targeted	Full Sample	Not Targeted	Targeted
API Standard	0.007	-0.006	0.015	0.208	0.629	-0.494	0.484	0.514	0.319
	(0.032)	(0.045)	(0.041)	(0.295)	(0.415)	(0.471)	(0.394)	(0.524)	(0.503)
API Plus	-0.026	-0.078*	0.007	-0.182	0.362	-0.758*	-0.043	0.028	0.038
	(0.035)	(0.045)	(0.046)	(0.319)	(0.419)	(0.387)	(0.362)	(0.406)	(0.538)
H0: Standard+Plus=	0.648	0.189	0.937	0.526	0.300	0.148	0.421	0.610	0.794
H0: Standard=Plus	0.371	0.144	0.874	0.261	0.575	0.554	0.247	0.403	0.606
Mean Dep. Var.	0.705	0.713	0.706	3.521	2.986	4.185	3.055	2.811	3.217
SD Dep. Var.	0.456	0.453	0.456	4.118	3.331	4.939	3.991	3.619	4.144
Adjusted R-squared	0.097	0.096	0.090	0.020	0.007	0.040	0.017	0.018	-0.000
Observations	1868	815	955	1176	506	608	1017	446	525
Clusters	224	179	218	215	161	203	206	150	197