IDB WORKING PAPER SERIES Nº IDB-WP-01427

Remote Tutoring with Low-tech means to Accelerate Learning: Evidence for El Salvador

Pablo Zoido
Iván Flores
Felipe J. Hevia
Miguel Székely
Eleno Castro

Inter-American Development Bank
Education Division

December 2022

IDB
Inter-American
Development Bank

Remote Tutoring with Low-tech means to Accelerate Learning: Evidence for El Salvador

Pablo Zoido
Iván Flores
Felipe J. Hevia
Miguel Székely
Eleno Castro

December 2022

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Remote tutoring with low-tech means to accelerate learning: evidence for El Salvador / Pablo Zoido, Iván Flores, Felipe J. Hevia, Miguel Székely, Eleno Castro.
p. cm. — (IDB Workin Paper Series ; 1427)
Includes bibliographical references.
1. Educational innovations-El Salvador. 2. Educational equalization-El Salvador. 3. Mathematics-Study and teaching-El Salvador. I. Zoido-Lobatón, Pablo. II. Flores, Iván. III. Hevia, Felipe. IV. Székely, Miguel. V. Castro, Eleno. VI. Inter-american Development Bank. Education Division. VII. Series.
IDB-WP-1427

JEL classification: I20, I24, I25, I28
Keywords: education research, inequality, economic development, government policy, education research

http://www.iadb.org

Copyright © 2022 Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode) and may be reproduced with attribution to the IDB and for any non-commercial purpose, as provided below. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Following a peer review process, and with previous written consent by the Inter-American Development Bank (IDB), a revised version of this work may also be reproduced in any academic journal, including those indexed by the American Economic Association's EconLit, provided that the IDB is credited and that the author(s) receive no income from the publication. Therefore, the restriction to receive income from such publication shall only extend to the publication's author(s). With regard to such restriction, in case of any inconsistency between the Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives license and these statements, the latter shall prevail.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

Remote Tutoring with Low-tech means to Accelerate Learning: Evidence for El Salvador

Pablo Zoido*, Iván Flores+, Felipe J. Hevia++, Miguel Székely+ and Eleno Castro* [1,2]

Summary

This paper presents the results of an impact evaluation, with an experimental design, which estimates the effect on learning math of a remote tutoring program offered to girls and boys aged 9-14 years in three departments of El Salvador. The program used low-tech interventions such as text messages and 20-minute phone calls over an eight-week period. Remote tutoring is estimated to have had a positive and significant effect of 0.23 standard deviations, which is equivalent to a 33.2 percent acceleration in learning math, compared to the changes observed in the control group. Evidence shows that the rate of learning increases considerably as the number of tutoring sessions taken increases. When compared with other related studies, we conclude that the intervention is cost-effective. The main innovative elements are: (i) the generation of evidence through instruments applied in person, which ensures high quality and accuracy in learning level measurements; (ii) to our knowledge, this is the first experimental evaluation program of its kind in Latin America implemented during the pandemic, with schools partially open, which allows us to verify whether the intervention is as effective as similar such interventions in other contexts and regions of the world; (iii) the use of two types of learning level tests to validate the robustness of the results.

JEL classification: I20, I24, I25, I28
Keywords: education research, inequality, economic development, government policy, education research

December 9, 2022

[1] *Inter-American Development Bank (IDB); +Centro de Estudios Educativos y Sociales (Center for Educational and Social Studies, CEES); ++Centro de Investigaciones y Estudios Superiores en Antropología Social (Center for Research and Higher Studies in Social Anthropology, CIESAS).

[2] The authors would like to thank Gonzalo Almeyda, Jenny Carolina Hernández, Victoria Oubiña and Alan Picazo for their help in the preparation of this study, as well as Samuel Berlinski, Matías Busso, and Gregory Elacqua of the IDB for their relevant and detailed comments on previous versions of this document. The authors would also like to express their gratitude to Youth Impact (Young 1ove as it was formerly known), in particular Noam Angrist and Colin Crossley, who shared the materials they had worked on in other regions of the world and offered technical support and advice in contextualizing them for El Salvador. The opinions expressed in this study are those of the authors and do not necessarily reflect those of the institutions to which they belong.

Contents

Introduction

The Covid-19 pandemic has had a significant impact on education systems around the world. Latin America and the Caribbean (LAC) was no exception, with some 114 million students dropping out of face-to-face classes in 2020. In 2021, the region ranked as the one with the highest number of affected students worldwide. The return to school was gradual, and by mid-2021, 86 million children in LAC had not yet returned to face-to-face classes (UNICEF 2021a, 2021b).

The interruption of traditional education mainly affected the poorest population as they have fewer resources for effective distance education, such as having access to internet at home, and having the necessary infrastructure such as computers or tablets and adequate physical space for remote study, the situation being similar in the schools in which these children study. The effects of the pandemic in the region included a decrease in the number of hours dedicated to study by students, a reduction in the number of hours offered by teachers, and less interaction of students with their peers and teachers (Acevedo et al., 2022).

These interruptions in the educational process have had considerable consequences on learning, some of which have been measured in recent studies. For example, in the State of Sao Paulo in Brazil, based on the standardized test for elementary education (SAEB) implemented in 2021, losses of between 11 and 29 percentage points were observed in Portuguese language scores and between 18 and 47 percentage points in math scores (Lichand, et al., 2021). In Colombia, the results of the *Plan de Alternancia* (Alternation Scheme) program, whereby some schools were selected for a progressive return to face-to-face classes at the end of 2020, showed that those children in schools that did not participate obtained lower results in math, reading and natural sciences tests (Melo-Becerra et al., 2021). In Chile, the Ministry of Education (MINEDUC, 2021) conducted a study comparing the 2020 DIA test results with those obtained prior to the pandemic, showing a reduction of between 40 and 53 percentage points in reading comprehension and math scores, respectively.

For Mexico, Székely et al. (2022) using data from the MIA learning test for the states of Yucatán and Campeche, show that children between 10 and 15 years old who had not returned to face-to-face classes experienced a learning loss of 0.28 standard deviations over a period of six months, which mainly affected older children. This information is consistent with that reported by Hevia et al. (2022a) for a similar age population. Finally, by means of a simulation, Monroy-Gómez-Franco et al. (2022) also estimate that in the case of Mexico a current learning loss equivalent to one third of an academic year can have an impact of between 1 and 2.1 years of learning loss in the long term, affecting mainly low-income children.

In El Salvador, the effects of the pandemic on school-age children have also been documented. Regarding learning processes, a study by Fundación Pro-Educación de El Salvador (El Salvador Pro-Education Foundation, FUNPRES) in 2021 for the national Ministry of Education in which questions were asked to a representative group of the school population, found that 58.5 percent considered that it was difficult to take online classes, 30.1 percent rated their access to the Internet as unfavorable and 19.8 percent stated that they had

inadequate mobile devices.

This same study revealed that there were higher levels of depression and anxiety than before the pandemic, affecting women and older children the most. Some of these problems were expected to diminish with the return to school. Although the process of reopening schools in El Salvador began in April 2021, by the middle of that year only 44.5 percent reported that they attended class three or more days a week, while 26.7 percent did not attend a single day.

In this context, this study reports the results of a low-tech remote tutoring intervention to accelerate learning in El Salvador. To our knowledge, this is the first intervention involving experimental evaluation of its kind in Latin America carried out during the pandemic. It was aimed at children from 9 to 14 years of age in the fourth to ninth grades of elementary education, in the departments of Chalatenango, Cabañas and Morazán. The intervention consisted of tutoring appropriate to the math level of each child.

The evaluation has an experimental design. The first step consisted in a diagnostic questionnaire (baseline) being taken in person by all participants, to ensure that the information captured in the learning measurements is reliable. Based on the results, a personalized intervention strategy tailored to the needs of each participant was implemented. The intervention consisted of personalized text messages and 20-minute phone calls for individual tutoring for 8 weeks. The results were measured by a face-to-face questionnaire at the end of the intervention. The participants took two types of tests to ensure the robustness of the results. We found that remote tutoring had an impact of 0.23 standard deviations, which is equivalent to a 33.2 percent acceleration in learning math, compared to the changes observed in the control group. The results suggest that the level of learning increases when the number of tutoring sessions taken is higher.

The document is structured as follows. The first section provides a brief review of the literature related to the study, and also describes the program and its experimental design. The second section presents the results of the fieldwork and basic descriptive statistics, while the third outlines the empirical strategy. The results of the impact assessment are presented in the fourth section and finally the conclusions are presented.

1. Program description and experimental design

Literature review

In order to increase educational achievement and mitigate the effects on learning, one of the interventions that has proven to be successful, mainly in children who show delays in their learning according to the corresponding curriculum, has been academic tutoring. These interventions may vary in terms of the type of tutors (teachers, semi-professionals, volunteers, peers, parents), presence (in person or remote), location (at home, at school after school, during class time), one-on-one or in small groups, etc. Tutoring, in general, is not a recently designed intervention and its positive effects have already been documented. In a meta-analysis of 96 post-1980 publications of randomly assigned tutoring interventions (and excluding those where tutors were peers) Nickow et al. (2020) show that these collectively have an impact of 0.37 standard deviations on learning variables, equivalent to moving from the 50th percentile to the 66th percentile of the distribution. Three years earlier, in another systematic review and meta-analysis, Dietrichson et al. (2017) considered 101 interventions in Europe and OECD countries (randomized and non-randomized) in students with low socioeconomic status, finding that tutoring had a pooled effect of 0.36 standard deviations.

Based on previous experience, remote tutoring has recently been used as a policy tool specifically to lessen the impact of the pandemic on learning loss. This is due to the fact that among its main characteristics are the fact that it is low-cost and uses the infrastructure available in homes, which makes it possible to serve children of different socioeconomic statuses, including the lowest. Hevia et al. (2022) present a literature review as well as a systematization of remote tutoring programs as an intervention to accelerate learning and reduce existing gaps. Particular examples of such programs include the pilot Tutoring Online Program (TOP) of one-on-one online tutoring for sixth through eighth graders, with college-level students acting as volunteer tutors. The intervention consisted of three hours of remote (online) tutoring per week for five weeks, with an intensive version of the program of six hours per week. This experimental design program was carried out at the beginning of the pandemic, between April and June 2020, in Italian high schools. Carlana et al. (2021) find that the intervention had an effect of 0.26 standard deviations on the score of a test that included math, Italian and English, with greater effects on lower-income and migrant children. They also demonstrate that there are significant effects on socioemotional skills and aspirations regarding studies. In Spain, the *Menttores* (Mentors) program served students in first and second grade of secondary education between 12 and 15 years of age, randomly assigned tutoring sessions (one tutor for two students), in a video-call format in 50-minute sessions three times a week for eight weeks and where the tutors were mostly professionally qualified teachers.

The program had a positive effect of 0.26 standard deviations on standardized math tests, positive effects on end-of-year math scores, and on the reduction of the probability of repeating the school year (Gortazar et al. 2021).

Regarding low-tech remote tutoring programs using phone calls and text messaging in the context of the pandemic, there is also positive evidence. In Bangladesh, a telephone tutoring program was carried out among elementary school children. The tutoring sessions were randomly assigned and the program consisted of a total of 13 weekly 30-minute sessions in math and English. Tutoring was also provided by telephone to mothers, supplemented by text messages to help them improve their learning at home. Effects of 0.56 and 0.66 standard deviations were found in math and English respectively as well as an increase of 0.64 standard deviations in the time parents spent helping their children with schoolwork (Hassan et al. 2021).

A similar intervention was recently conducted in Botswana with children in grades three through five of elementary school, in which the tutoring was randomly assigned in two forms. The first consisted of text messages with math problems, and the second included weekly calls of 15 to 20 minutes for a total of three hours over eight weeks. The improvement in math learning in the combined message and call tutoring treatment was 0.12 standard deviations, also finding a 31 percent reduction in the number of children who could not perform math operations (Angrist et al. 2022).

Finally, another low-tech intervention was conducted in Nepal where 3,700 households of public-school children in grades three through five of elementary education were randomly assigned different treatments. It combined text messages and calls from teacher tutors or facilitator tutors from social organizations. In this case, effects of up to 0.19 standard deviations are found in math learning, representing a 30 percent increase with respect to the control group (Radhakrishnan et al. 2021).

Description of the intervention

The Remote Tutoring Program to Accelerate Learning in El Salvador is an initiative of the Inter-American Development Bank in conjunction with the Ministry of Education, Science and Technology, based on previous interventions such as those carried out in Botswana and Nepal (Angrist, Bergman & Brewster, 2022), which aims to accelerate learning, in the context of the pandemic, through the use of low-tech platforms (that do not require internet or hardware inputs such as laptops or tablets). This program is expected to reduce learning gaps among students from the most vulnerable sectors by delivering a series of eight personalized telephone tutoring sessions to students between the ages of 9 and 14, lasting approximately 20 minutes each, working on specific mathematical operations in each of these calls.

Due to its characteristics in terms of duration and focus on math learning, this intervention was planned not as a substitute but as a supplement to formal education (face-to-face, distance or hybrid).

The intervention consisted of three phases: planning, tutor recruitment and training, and tutoring. First of all, in the planning phase, students between 9 and 14 years of age were identified as the target population, considering the transition period from elementary to secondary school, and that in the pandemic the situation of "learning poverty" in math would have increased in this population. Second, the pedagogical strategies were defined with this population. Based on the existing literature, three principles were defined: 1) focus on basic skills; 2) simplicity in its implementation and scaling; and 3) personalization, recognizing the needs of each student. Therefore, we chose to use the pedagogical strategy of "Teaching at the Right Level" (TaRL) which involves identifying the actual capabilities that each student has and placing them within different levels of difficulty (Angrist, et. al., 2020; Banarjee et. al., 2016; Karthik, Singh & Gainimian, 2017). Third, the tutoring sessions' contents and levels of difficulty were selected, seeking to align them with the national curriculum. In this intervention, five levels were identified: place value, addition, subtraction, multiplication and division, which in the national curriculum correspond to operations that should be ensured between the 1st and 4th grades (Annex A1). With this information, a program of eight tutoring sessions each lasting 20 minutes, delivered over eight weeks, was designed and intervention guides for the tutors were developed. Finally, a tutor profile was defined based on two criteria: experience in teaching math and interest in participating in this project.

The second phase consisted of recruiting and training tutors. To this end, tutor recruitment campaigns were carried out with the support of the Ministry of Education. Interested tutors enrolled in the program and each received three items of support: 1) 8 hours of synchronous remote training, with 50 percent of the time spent reviewing practical cases and exercising the tutoring function; 2) a written guide explaining the operation of the project, and the activities that each tutor had to follow, sensitizing parents to participate, explaining the mathematical processes to the children and filling out the daily log, and; 3) remote accompaniment (through messaging groups) by a team of specialized coordinators to resolve doubts and to follow up on the development of the tutees. A total of 56 tutors participated in this intervention, 74 percent of whom were women, 50 percent were classroom teachers, 12 percent were student teachers, 16 percent were university students and 20 percent were professionals engaging in other activities who met the profile.

The third phase was delivery of the tutoring sessions. In the first stage, each student was given two learning measurement instruments to identify the appropriate level, the IDB-Young 1ove test and the MIA+ (MIA Plus) test. Thus, children who could add, but could not subtract, started at the "addition" level, while children who could multiply, but could not divide, started at the "multiplication" level.

In this way, tutoring sessions focused on a specific operation, and only moved on to the next operation when the student demonstrated mastery in a series of challenges.

Once the tutors were trained and the appropriate level of each participant was identified, awareness calls to the families in the treatment group were made to explain the program to the parents. If they agreed to participate, a second confirmation call was made where a tutor was randomly assigned to organize the day and time of the tutoring session. At the time of receiving the telephone calls for tutoring, we requested the presence of an adult or responsible person to accompany the student in this process. Thus, the parents' participation in this program consisted of giving their approval for the inclusion of the children and accompanying the tutoring process.

Each of the eight tutoring sessions followed a similar structure. First, the results of the operations and problems of the level in which each student was, and which had been sent in advance by SMS, were reviewed, resolving any doubts and problems that the student might have. Second, new problems of the same level of difficulty and more exercises to solve the operations were presented, and a step-by-step explanation of the procedures to solve the mathematical operations was given. Third, a series of challenges in the form of everyday problems were applied in order to confirm the student's mastery of the operation. If the student answered the SMS, exercises and challenges correctly, the next tutoring session started with a new operation. At all times, the tutor gave feedback on the procedures and answered any doubts students might have. And at the end the tutor gave feedback on the session to the students. Once the call was over, the tutors filled out a daily online follow-up log to monitor each participant's performance week by week. At the end of the eight tutoring sessions, parents, students and tutors completed a satisfaction survey and two focus group sessions were held with the tutors to systematize the experience.

The theory of change of this intervention establishes that once the inputs are identified (learning to be measured, target population with learning lags, adequate pedagogical guidance and the corresponding training of tutors), regular and effective communication processes for teaching-learning will be carried out between tutors and the children of the target population, through personalized follow-up, generating a relationship of trust: a "mentoring relationship". These processes will be carried out using media within the reach of the target population and will lead to improvement in fundamental math learning, as well as tutors having self-efficacy perception, and satisfied students, tutors and parents.

Due to its characteristics of using media such as telephones and text messages, this program could be implemented regardless of the modality in which the children were taking their classes: in the baseline surveyed at the end of 2021, 86.6 percent of the children reported that they attended classes in person, 2.4 percent were taking distance classes and 11 percent were taking hybrid classes.

Experimental design

The intervention was carried out in three departments of El Salvador: Cabañas, Chalatenango, and Morazán. The focus on these departments was based on discussions with the Ministry of Education. These departments are characterized by being border departments, having poverty levels above the national average, having lower literacy rates and lower school attendance rates according to the General Directorate of Statistics and Census (DYGESTIC, 2022), statistics shown in Table A2 of the Annex. The target population was students in grades 4 through 9 of elementary education, between 9 and 14 years of age. The children attending these grades in the three departments total 5,567 students.

The question to be answered in this study was: what is the effect of tutoring on math learning? Taking this into account, a minimum detectable effect of 0.13 standard deviations was defined for the learning test (lower than that reported in the studies cited in the previous literature review), a statistical power of 80 percent and a confidence level of 95 percent, which determined a necessary sample size of 1,830 children interviewed. Considering two survey rounds and a sample loss of more than 40 percent, it was estimated that around 3,200 interviews should be planned at baseline.

In practice, a higher number of interviews was achieved, the total standing at 3,440. Randomization was performed among the children who answered the baseline questionnaire, at the individual level and without stratification, leaving 1,720 assigned to the remote tutoring program and 1,720 to the control group. The balance tests show that both groups (treatment and control) have a highly similar profile in all relevant characteristics measured prior to treatment, including initial math level, attendance modality (face-to-face or remote), enjoyment of studying. The balance tests are presented in Table A3 in the Annex.

Impact measurement tools

We collected data in schools. This consisted of the face-to-face implementation of a baseline questionnaire and a post-tutoring follow-up questionnaire designed specifically for the experiment. The items included in the instrument were the general socioeconomic status of the students, their family socioeconomic status, the existence of a disability or special needs, the level of learning in math, as well as items to measure mental health and the degree of motivation of the students.

To measure learning, we used two instruments, which underwent standard content and reliability validation processes (León et. al., 2022). The first was IDB-Young 1ove, which consists of nine items that measure place value, addition of two digits with carry-over, subtraction of two digits with carry-over, multiplication of two digits by two digits, division of two digits by one digit with a remainder, a mathematical problem, two logical problems, one of which was optional, and a sum of fractions. For each item only one possible answer was given and had to be answered in less than 2 minutes each. The first five items (place value, addition, subtraction, multiplication, division) were used to identify the appropriate level. All items were applied one by one and achievement-error responses were given. This instrument has an adequate internal discrimination (item-retest correlations greater than 0.20 in all cases), an adequate internal consistency index (Cronbach 0.766 and Omega 0.778), with unidimensional construct validity.

The second instrument was MIA Plus. This instrument measures basic mathematical operations and problem solving. It consists of nine items: number, addition of two digits without carry-over, addition of two digits with carry-over, subtraction of two digits without transformation, subtraction of two digits with transformation, division of three digits by one digit without remainder, addition problems with visual support, multiplication problem without visual support, and addition of fractions with a different denominator. In the case of the operations items, subjects had to answer at least two out of three options correctly, and for the problem items the subject could answer correctly up to two times. The first six items (number, addition one, addition two, remainder to one, subtraction two, division) were used to identify the appropriate level for tutoring. This instrument applied the items by level of difficulty and discontinued its application when the subject was unable to answer one of the items. This instrument had an adequate internal discrimination (item-retest correlations greater than 0.20 in all cases), an adequate internal consistency index (Cronbach: 0.81 and Omega 0.88), with a three-factor structure with adequate fit indicators.

Mental health

Finally, in order to obtain an approximation of the mental health status of girls and boys in the questionnaire, we included questions from an instrument validated in a population similar to that of this study (Hernández-Guzmán et al., 2010), the Spence Children's Anxiety Scale (SCAS). Children in El Salvador had different levels of anxiety and depression than those measured prior to the pandemic according to FUNPRES (2021), so having this information allows these factors to be considered in the program's impact measurement estimates.

The SCAS consists of 38 questions for anxiety issues, prepared for children to answer on a Likert-type scale: never (0), sometimes (1), many times (2) and always (3). The questions are grouped into six subscales which are: 1) separation anxiety, 2) social phobia, 3) obsessive-compulsive disorder, 4) panic disorder/agoraphobia, 5) fear of physical injury, and 6) generalized anxiety. In accordance with the methodological recommendation for the interpretation of the SCAS results, we used the T-Scores method, which takes into account the age and sex of the individuals, following the pre-established cut-off points for the interpretation of the results. The latter are based on the scores corresponding to each percentile of the distribution of each subscale in a reference population; the scores vary according to the age and sex of the population and the subscale considered.

2. Fieldwork results and descriptive statistics

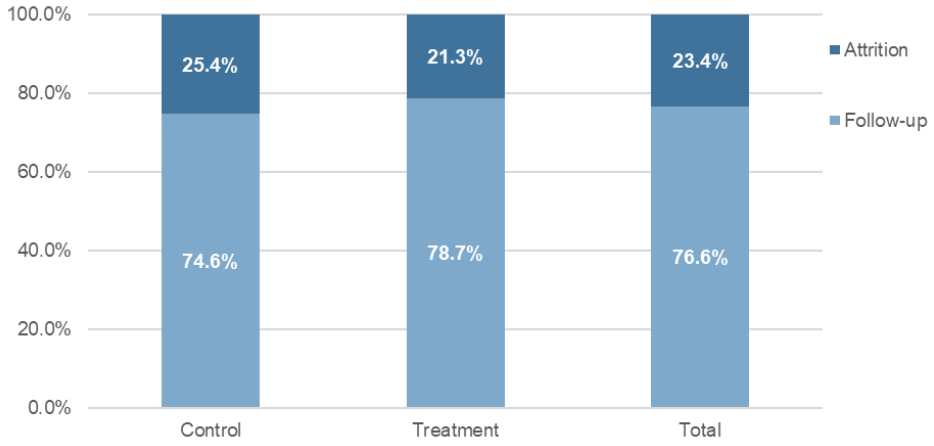
We gathered information for the baseline in the months of October and November 2021, and conducted 3,440 interviews, in accordance with the planning. After this first stage, a total of 810 children took at least one tutoring session. The tutoring sessions were given by 59 tutors, of whom 76 percent were women, 46 percent had bachelor's degrees and 56 percent were teachers. Their average age was 31 years, with a range from 19 to 63.

Follow-up and Attrition

In the post-intervention follow-up round, 2,636 students were interviewed, equivalent to 76.6 percent of the baseline population. It is important to note, however, that the final sample size was larger than planned based on power calculations.

The main reason why some interviews were not successful was due to communication problems, where the family was not found or did not receive an answer via telephone, partly because of the lack of security in the country, which can make families distrustful of telephone calls from numbers they do not recognize. The second reason was that it was not possible to obtain parental consent for their children to participate in the follow-up, which is why we did not continue with this stage in these cases.

Figure 1. Percentage of participation in the follow-up



In practice this implied that the total attrition was 23.4 percent, with a differential of 4.1 points between the control group (25.4 percent attrition - 437 observations) and the treatment group (21.3 percent - 367 observations). Therefore, the final sample size was 1,283 and 1,353 participants, respectively.

The combination of 4.1 points of differential attrition with a total attrition of 23.4 percent is considered low potential bias according to standards published by *What Works Clearinghouse* (WWC, 2022) of the U.S. Institute of Education Sciences, who group the percentage of attrition in experimental design evaluations into low or high, relative to the level of potential bias.[3] These categories vary in relation to the total percentage of attrition and differential attrition. However, to establish whether the treatment and control groups differed in their characteristics in the follow-up, due to attrition, a mean difference test was performed for girls and boys who participated in the follow-up.

Table 1 shows the means and standard errors of the characteristics of the girls and boys who answered the follow-up questionnaire, as well as the differences between groups: no differences are observed between children in the treatment and in the control that are significant at 5 percent, so both groups are similar in terms of characteristics measured at baseline such as sex, age, disability, liking for study, classroom attendance, socioeconomic status, anxiety levels, math level, and department of origin. Table A4 in the Annex presents the differences between children in the treatment and control sample that did not participate in the follow-up, with similar results.

[3]A study is defined as having low attrition if its potential bias is less than 0.05 standard deviations, according to the algorithm described in WWC (2022) and in accordance with the values shown in Table C.1 of that manual.

Table 1. Balance of (baseline) characteristics in the follow-up sample

	Control	Treatment	Difference
Female	0.525	0.514	0.012
	(0.014)	(0.014)	(0.0195)
Age	11,799	11,718	0.081
	(0.043)	(0.044)	(0.0619)
Age - 9 to 11	0.426	0.46	-0.033*
	(0.014)	(0.014)	(0.0194)
Disability	0.051	0.056	-0.005
	(0.006)	(0.006)	(0.0088)
Grades 7 to 9	0.276	0.285	-0.009
	(0.012)	(0.012)	(0.0175)
Face-to-face	0.878	0.858	0.020
	(0.009)	(0.009)	(0.0132)
Enjoys studying a lot	0.844	0.846	-0.002
	(0.010)	(0.010)	(0.0141)
A great desire to learn	0.923	0.913	0.010
	(0.007)	(0.008)	(0.0107)
Severe over-age	0.052	0.044	0.008
	(0.006)	(0.006)	(0.0083)
Parents who did not graduate high school	0.684	0.678	0.006
	(0.013)	(0.013)	(0.0182)
SES Score	112,772	112,641	0.131
	(1,248)	(1,204)	(1.7337)
High SES	0.129	0.121	0.007
	(0.009)	(0.009)	(0.0129)
Medium SES	0.295	0.310	-0.016
	(0.013)	(0.013)	(0.0179)
Low SES	0.577	0.568	0.008
	(0.014)	(0.013)	(0.0193)
Generalized Anxiety	0.323	0.325	-0.002
	(0.013)	(0.013)	(0.0182)
Obsessive-Compulsive Disorder	0.355	0.363	-0.008
	(0.013)	(0.013)	(0.0187)
Social Phobia	0.276	0.276	0.000
	(0.012)	(0.012)	(0.0174)
Separation Anxiety	0.575	0.582	-0.006
	(0.014)	(0.013)	(0.0192)
Panic disorder/Agoraphobia	0.405	0.397	0.008
	(0.014)	(0.013)	(0.0191)
Fear of Injury	0.309	0.295	0.015
	(0.013)	(0.012)	(0.0179)
Chalatenango	0.280	0.274	0.006
	(0.013)	(0.012)	(0.0174)
Cabañas	0.451	0.452	-0.002
	(0.014)	(0.014)	(0.0194)
Morazán	0.270	0.273	-0.004
	(0.012)	(0.012)	(0.0173)
Correct Answers (%) IDB-Young	0.315	0.299	0.0158
	(0.008)	(0.008)	(0.0119)
Correct Answers (%) MIA Plus	0.353	0.348	0.005
	(0.006)	(0.006)	(0.0089)
Observations	1.283	1.353	

Notes: Standard errors in parentheses. Levels of significance: * p< 0.10, ** p < 0.05, *** p < 0.01

A comparison, characteristic by characteristic, between the sample that did participate in the follow-up and the sample that did not participate is shown in Table A5 in the Annex, where it is observed that those who were interviewed were 0.2 years older, suffered slightly more anxiety, belonged less to the department of Chalatenango and had a higher level of math at

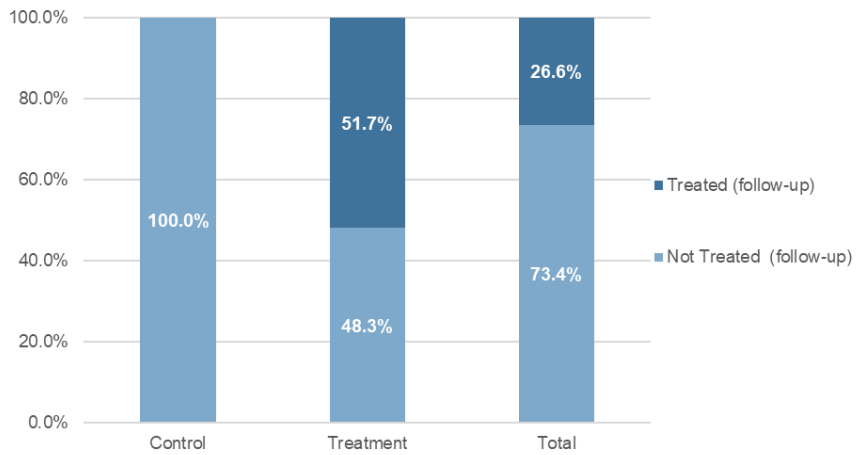
baseline, which marks some differences with respect to the initial population.

On the other hand, to explore a possible selective attrition (to know if these differences are related to having been assigned to control or treatment), we performed a joint test consisting of a regression in which the dichotomous variable to be explained was attrition and the explanatory variables were a) the intention-to-treat, b) control variables and c) the interaction of the intention-to-treat variable with each of these control variables. The results are presented in Table A6 of the Annex. Similar to what is observed in Table A5, attrition was mainly associated with the department, with Chalatenango having the highest attrition and levels of anxiety in children, which were positively associated with attrition. However, only the interaction of the intention-to-treat with separation anxiety was significant, standing at 5 percent, and there was no association between characteristics that selectively affected attrition and also one of the treatment groups. This is reflected, for example, in Table 1, in which there are no significant differences, at 5 percent, in the characteristics observed between treatments and control groups that answered the follow-up questionnaire.

Non-compliance

Of the 1,283 children participating in the follow-up who were randomly assigned to the control group, as expected, none were treated (Figure 2), i.e. there was no non-compliance with assignment to the control group. In the implementation of the program, no untreated children were known to show interest in participating or taking tutoring sessions. On the other hand, of the 1,353 children in the treatment group found for the follow-up, 700 took at least one tutoring session, or 51.7 percent of that group. Of these, 81.7 percent took a total of eight tutoring sessions and only 10.1 percent took only one or two.

Figure 2. Percentage of children who took tutoring (treated) with respect to the original intention-to-treat, sample of children participating in the follow-up



The main reason for non-compliance in the treatment group was problems establishing contact prior to the start of the tutoring preparation process, due to unanswered calls. The second reason was a refusal to participate or an initial acceptance that was never reached. Despite this, Table A7 in the Annex presents balance tests that corroborate that the observable

characteristics of those assigned to the treatment and those that were treated and untreated, respectively, are similar.

Factors associated with the conclusion of the Program

Additionally, we made a comparison between the treated children, but distinguishing those who took the planned number of tutoring sessions (8 tutoring sessions) with those who did take the treatment but to a lesser extent: a) 1 or 2 tutoring sessions and b) 3 to 7 tutoring sessions, a comparison shown in Table A8. The children who completed the tutoring sessions are very similar to those who interrupted them at the beginning (1 or 2 tutoring sessions) in terms of characteristics observed, with the only significant difference at 1 percent being the percentage of children with a disability, which was higher in the latter group. Contrary to expectations, there were more differences between children who completed the program and those who took between 3 and 7 tutoring sessions, for example in relation to age for school grade and liking for study, suggesting that non-completion of the program is not a purely random phenomenon. This should be taken into account in future implementations in order to maximize the number of tutoring sessions children take in this intervention. In terms of what was observed in the field, of the total number of children who took at least one tutoring session only 18.3 percent did not complete the 8 tutoring sessions and the main reason was that, of these, three quarters stopped answering the calls and the cause could not be ascertained.

3. **Empirical strategy**

To measure the impact of tutoring on children's learning, we use the following linear regression model:

$$(1) \qquad Y_i = \beta_0 + \delta D_i + \alpha X_i + \varepsilon_i$$

Where Y_i is a variable measuring learning in math in the follow-up for student i, while D_i is a dichotomous variable denoting whether child i was randomly assigned to the treatment. The vector X_i is individual, educational, socioemotional, family, geographic context, and corresponding math test score characteristics, all measured at baseline. With an experimental design, the coefficient δ is the estimated ITT (Intention-to-Treat) effect. The ITT is the average effect of the intention-to-treat on learning, regardless of whether the children assigned to receive treatment ultimately took the treatment or not. Determining this effect is useful because it allows us to determine the impact of replicating this program, offering it on a voluntary basis, in a similar population and with the same compliance rate.

In addition, we estimated the TOT (Treatment on the Treated) effect, which quantifies the program's impact on the children who actually took the tutoring. This effect is therefore greater because it is not "diluted" by the results of those children who were assigned to the

treatment but did not take it. The TOT is then the impact on those treated, which as shown in Table A7, in this case is a very similar population to those assigned to the treatment but who did not take the tutoring sessions. For this purpose, in an initial stage (equation 2) we estimated the value of variable T_i (indicating whether students took at least one tutoring session) as a function of the control variables and the intention-to-treat variable D_i which functions as an instrument (linear probability model). In a second stage we also estimated by OLS learning Y_i as a function of the control variables and the estimated $\widehat{T}_i$ whose coefficient ρ is the TOT.

$$T_i = \pi_0 + \gamma D_i + \pi X_i + v_i \quad (2)$$

$$Y_i = \alpha_0 + \rho \widehat{T}_i \, \alpha X_i + \omega_i \quad (3)$$

Finally, for each effect and for each test, the estimates considered two different specifications. In the first case, we did not include any control variables, which implies results equivalent to a comparison of means (in the case of the ITT), while in the second case they were included. This second model controls for observed characteristics that could influence the outcome, and therefore increases the precision of the estimated effects.

Learning variables

Estimates were made for the math learning variables using the two instruments included in the questionnaire: the IDB-Young 1ove test and the MIA Plus. At first, continuous variables were defined as the percentage of correct answers in each test. For IDB-Young 1ove, we used the first six items, leaving out those after the children are asked if they want to continue with the test (item eight). For the MIA Plus instrument, all questions were included. Second, we subtracted these variables from the mean percentage of correct answers in the follow-up and divided by the observed standard deviation of the controls. Finally we used the variables Y_i in units of standard deviation of the latter.

Control variables

The control variables included in the models were selected to represent the different factors that could influence children's learning. Although the questionnaire made it possible to collect a good amount of information, we used the most representative variables to capture the effects of the different factors.

Thus, as individual variables, we included sex, disability and age. As for school variables, we included dichotomous variables of school cycle[4], school attendance[5] and a motivation variable that indicates whether the child responded that they enjoyed studying a lot. Regarding socioeconomic status, an index (continuous variable) was included, taking into account the parents' education, the number of complete bathrooms and the number of rooms in the home, the number of cars, access to internet services and the number of people over 14 years of age who worked in the last month[6]. In relation to the socioemotional variables of the SCAS scale, we included the most prevalent among the population of children, which is separation anxiety, included as a dichotomous variable of high or very high level of anxiety. Differences in the local context were captured by including two dummy variables of the department where the children lived. Finally, the most important control variables that were included in the model are percentage of success variables in the respective math tests at baseline (IDB - Young 1ove or MIA Plus test scores, respectively).

4. **Results**

The main results are presented in Table 2 and show the positive and significant effects of tutoring on learning. The impact of the intention-to-treat (ITT) on girls' and boys' learning is 0.11 standard deviations for the IDB-Young 1ove test. This result is of a significant magnitude, considering that 51.7 percent of those assigned to receive treatment actually took at least one tutoring session. The effect of treatment on the treated (TOT) is about twice as big, standing at 0.22 standard deviations, also significant at 1 percent. The difference in size between these two effects is due precisely to the fact that a large percentage of those assigned to receive treatment were not actually treated.

[4] An indicator variable was included for grades seven through nine, which are the grades prior to the start of the last two years of high school. The students who participated could also be studying from fourth to sixth grade.
[5] Face-to-face attendance was defined as indicating the first response option to the questionnaire question "How do you take your classes?": a) I attend school, b) I take online classes, or c) both.
[6] This index originally developed for Mexico was generated in accordance with AMAI (2020) and is used taking advantage of the availability of information in the baseline questionnaire.

Table 2. ITT and TOT impacts (standard deviations)

	(1) ITT *OLS*	(2) TOT *2SLS*
IDB-Young 1ove		
No controls (N=2,636)	**0.112*** **	**0.216*** **
	(0.0386)	(0.0743)
With controls (N=2,636)	**0.121*** **	**0.234*** **
	(0.0360)	(0.0694)
MIA Plus		
No controls (N=2,636)	**0.078** **	**0.151** **
	(0.0385)	(0.0744)
With controls (N=2,636)	**0.080** **	**0.156** **
	(0.0357)	(0.0691)

Notes: Standard errors in parentheses.
Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01.

In the second specification, with control variables included, the standard errors of the impacts are smaller and the magnitudes are larger. The estimated ITT is 0.12 and the TOT is 0.23 standard deviations using the IDB-Young 1ove test. The impacts were corroborated using the MIA Plus test, which has the important feature that once a child makes a mistake on a question, the test is stopped and the remaining items are not presented. In this case the impact is positive and significant with magnitude 0.08 standard deviations for intention to treat (ITT), and is also significant at 5 percent with a magnitude of 0.16 standard deviations for those who actually took the tutoring sessions (TOT).

The complete results, including the coefficients of the control variables, of the ITT estimates for both tests are presented in Figure A1 in the Annex. Beyond the impact estimates, when observing the coefficients of the characteristics considered, it is observed that being older and studying in the higher grades (seventh to ninth) are associated with higher learning in math. Furthermore, belonging to Chalatenango is associated with lower scores in the tests, as is having some type of disability. Finally, prior learning in math, measured at baseline, is the most important variable among the control variables.

Number of tutoring sessions

As we expected the number of tutoring sessions to influence the extent to which learning improves, we estimated the impact of the program differentiating by the number of sessions that the students held. We considered two alternatives; in the first, the treatment variable was divided into three groups: 0 tutoring sessions, 1 to 7 tutoring sessions, or 8 tutoring sessions, which were the groups into which the children were grouped based on descriptive statistics[7]; in the second, we considered the number of tutoring sessions as a continuous variable.

At a first level it consisted of a linear regression model explaining learning with the control variables as well as the tutoring variable. At a second level, the instrumental variables method was used in two stages: in the first stage, we estimated the continuous variable of tutoring using the control variables and the intention to treat as an instrument. In the second stage, we

regressed the learning variable against the control variables and the number of tutoring sessions estimated in the previous stage. This second method would be an approximation of the TOT impact of the number of tutoring sessions, although it should be noted that we did not randomize the number of tutoring sessions at the intervention design stage.

Table 3 shows that there is a significant difference between taking the total number of tutoring sessions and taking 7 or fewer, with the only statistically significant association being when the tutoring is complete, a pattern that is repeated for both tests used. In particular, the magnitude of the coefficient of 8 tutoring sessions for the IDB-Young 1ove test is 0.29 standard deviations. As for the continuous variable, the results using OLS and two-stage OLS do not differ much, being statistically significant at 1 percent for the aforementioned test. The magnitude of the coefficient is 0.033 standard deviations per tutoring session, which is consistent with the above estimate for the 8-tutoring session variable. The results suggest that there is a dose-effect relationship between tutoring sessions taken and learning math, although it should be considered that the dose was not randomized.[8]

[7] The percentages of children in the follow-up who took 0 tutoring sessions, 1 to 7, or 8 tutoring sessions were 73 percent, 5 percent, and 22 percent, respectively. In this case, the category *0 tutoring sessions* includes both those assigned to the control group and those assigned to the treatment who did not take any tutoring sessions.

[8] Similar to this exercise with the number of tutoring sessions, we performed an OLS regression where the treatment variable was divided into three groups according to the profile of the tutor: students, teachers and others. No statistically significant difference was found between the coefficients of each of these profiles.

Table 3. Learning in math by number of tutoring sessions (standard deviations)

	(1) OLS	(2) 2SLS
IDB-Young (N=2,636)		
In dichotomous variables		
1 to 7 tutoring sessions	**0.022**	
	(0.0793)	
8 tutoring sessions	**0.290*** **	
	(0.0429)	
Continuous variable	**0.036*** **	**0.033*** **
	(0.0053)	(0.0098)
MIA PLUS (N=2,636)		
In dichotomous variables		
1 to 7 tutoring sessions	**-0.051**	
	(0.0738)	
8 tutoring sessions	**0.170*** **	
	(0.0431)	
Continuous variable	**0.021*** **	**0.022*** **
	(0.0053)	(0.0097)

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01. Column (1) shows the results of four regressions, two for each test: one with two dichotomous variables of number of tutoring sessions taken and one with a continuous variable.

Subgroups

To explore whether tutoring generates a greater benefit by gender, age and in some groups of children more than in others, mainly in those with more unfavorable conditions, the empirical strategy for estimating the ITT was replicated but modified to include an interaction with one of the subgroups of interest.

The first group considered was lower-income children, for which the dichotomous variable of low socioeconomic status was constructed, where 56.9 percent of the baseline sample and 57.2 percent of the sample in the follow-up were located. This cutoff point chosen was a score below 115 points on the AMAI socioeconomic index, which is the cutoff point for a socioeconomic status close to the percentage of children living in households in poverty in the three departments considered in this study (DYGESTIC, 2022). The second group was generated considering the math level at baseline, used to assign children to tutoring according to their level: addition, subtraction, multiplication, or fractions. The low level of math was defined as those in the first category. Finally, three additional groups related to individual characteristics were considered: gender, age and having some disability.

Table 4. Learning in math by population subgroup (standard deviations)

	(1) *SocioE Status. Low (SES)*	(2) *Females*	(3) *Low level in math test (Baseline)*	(4) *Age 9 to 11*	(5) *Disability*
IDB-Young 1ove *(N=2,636)*					
ITT	**0.039**	**0.097***	**0.162*****	**0.087***	**0.129*****
	(0.052)	(0.052)	(0.058)	(0.047)	(0.037)
ITT*Var	**0.143****	**0.047**	**-0.071**	**0.071**	**-0.154**
	(0.072)	(0.072)	(0.074)	(0.073)	(0.166)
MIA PLUS *(N=2,636)*					
ITT	**0.036**	**0.061**	**0.189*****	**0.052**	**0.091*****
	(0.055)	(0.050)	(0.057)	(0.048)	(0.037)
ITT*Var	**0.078**	**0.036**	**-0.170****	**0.061**	**-0.194**
	(0.072)	(0.071)	(0.073)	(0.072)	(0.157)

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01. *Low level* identifies children with low performance on any of the Baseline math tests, a category that was used in assigning the level at which the child's tutoring would begin in the treatment. Each column (for each instrument) is the result of a separate regression, which includes control variables, the intention-to-treat and, additionally, an interaction of the treatment variable with the variable of interest, the coefficients of these two variables being those shown in the table.

In Table 4 each column represents a separate regression for the learning variables for each test, in standard deviations, and only the coefficients of the two treatment variables included, the intention-to-treat variable and the one interacted with the variable of interest, are presented. For the IDB-Young 1ove test, only the interaction of the treatment with the low socioeconomic status variable is positive and significant at 5 percent and with a magnitude of 0.14 standard deviations, which suggests that children of lower socioeconomic status take greater advantage of the tutoring in terms of their level of learning in math; however, in the MIA Plus test this result is not corroborated. In this last test the result is similar because there are no effects of the program for any specific group, except for the initial level of math, which is also significant at 5, with a negative sign and a magnitude of 0.17 standard deviations. This would suggest that tutoring has a greater impact on students who already have a better level of math. Again, this relationship is observed in only one of the two tests, the results are not conclusive[9].

[9] Although the number of tutoring sessions was not randomized in the intervention design, we replicated the subgroup analysis by replacing the intention-to-treat variable with the continuous variable of number of tutoring sessions. No consistent results were found with respect to differences by subgroups.

Robustness of results

As robustness tests, we replicated the calculations of program effects in both tests for the standardized score variable, as well as for the individual effort (optional logic) and fraction questions. The first method used was inverse probability weighting including the same control variables as for the main estimations. The second method was the Lee Bounds method, which instead of a point estimate provides a range of the program impact. The results are presented in Table 5. [10]

Table 5. Estimates using inverse probability weighting

	(1)	(2)	(3)	(4)
	Standardized Score		**Effort**	**Fractions**
	IDB-Young 1ove	*MIA Plus*		
Inverse Probability Weighting				
IPW	**0.122*****	**0.079****	**0.011**	**-0.004**
	(0.036)	(0.036)	(0.019)	(0.017)
Lee Bounds				
Low Level	**0.112*****	**-0.02**	**-0.019**	**-0.078****
	(0.048)	(0.053)	(0.023)	(0.046)
High Level	**0.212****	**0.179*****	**0.035**	**0.002**
	(0.054)	(0.050)	(0.022)	(0.033)
Observations	2,636	2,636	2,636	2,636

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01. Estimates for the dependent variables of standardized score of the indicated tests and for the Effort and Fractions questions of the *IDB-Young 1ov*e test as dichotomous correct or incorrect answer variables. The IPW method uses the same control variables as for the ITT and TOT estimates.

The inverse probability weighting results are almost identical to those obtained previously. On this occasion, an effect of 0.122 standard deviations (in contrast to 0.121 for the ITT estimate) is estimated for the IDB-Young test and 0.079 (in contrast to the 0.080 standard deviations obtained before) for the MIA Plus test. Finally, Lee Bounds estimates for the impact measured by the IDB-Young test are both positive and significant and suggest an impact of even greater magnitude. The lower bound for MIA Plus is not significantly different from zero, although the midpoint of the range coincides with the ITT estimate. The similarity with the main results provides robustness to the study's estimates.

[10] As an additional robustness test, we replicated the estimates with school fixed effects, with very similar results for the ITT and TOT. There were 46 schools where the students who answered the follow-up questionnaire were studying, although the initial randomization did not consider them as strata nor were they randomized in clusters, so we did not include them in the main model.

In addition, the IDB-Young 1ove questionnaire included a logic question characterized by requiring only minimal numerical skills to answer it correctly, so it would function more as a proxy for the effort to answer the questionnaire than for specific mathematical skills.[11] The results are shown in Table A9 of the Annex and it is not observed that tutoring is associated with a higher percentage of correct answers on this item. This suggests that the program does indeed result in better test performance due to improved cognitive skills and not due to increased test-taking effort. The table also shows the effect of the program on the last question of the questionnaire, which was a sum of fractions. The purpose of including this question was to find out if the tutoring generated improvements in learning beyond the planned agenda, which did not include this type of problem. The results were not positive or significant, so we do not have the elements to affirm that the program has positively impacted skills beyond those planned in the intervention.

Results in context

In both the control and treatment groups there was an improvement in learning between the two surveys, which is due in part to the increase in the proportion of children who returned to class in person[12]: at baseline 86.6 percent of the children mentioned attending in person, reaching 98.9 percent in the follow-up, with no statistically significant differences between treatments and controls in the percentages of in-person attendance. Taking this into account, an alternative way to understand program results is to present them in terms of the progress shown by the control group between the baseline and follow-up.

To this end, a variable was first generated as the difference between the results between the two survey rounds, and explained by linear regression using the controls from the previous models. The average progress of the controls, estimated by this regression, served to standardize the math score variables and obtain the main estimates again. The results are presented in Table 6.

[11] The contextual question was: "The day before the day after tomorrow is Saturday. What day is today?

[12] In recent studies in India and Mexico, Singh et al. (2022) and Székely et al. (2022), respectively, show that the reopening of schools has had a positive impact on the recovery of lost learning.

Table 6. Impacts in terms of progress of the control group

	OLS	*2SLS*
IDB-Young 10ve *(N =2,636)*		
No controls	**0.159*****	**0.306*****
	(0.055)	(0.106)
With controls	**0.171*****	**0.332*****
	(0.051)	(0.099)
MIA PLUS *(N =2,636)*		
No controls	**0.142****	**0.274****
	(0.070)	(0.135)
With controls	**0.146****	**0.283****
	(0.065)	(0.125)

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01.

The first column of Table 6 shows the ITT results for both tests. The impact of the intention-to-treat is 17.1 percent greater gains in math learning for the specification with controls on the IDB-Young 10ve test, while for the MIA Plus test this impact was 14.6 percent. On the other hand, the magnitude of the impact when considering only the population that did take the treatment (TOT) was 33.2 and 28.3 percent, for each test respectively, for the regression with controls.[13]

Table 7 shows the same exercise but considering the number of tutoring sessions instead of the main treatment variable (one or more tutoring sessions). For the IDB-Young 1ove instrument the progress was 41.3 percent for the children who completed the 8 tutoring sessions compared to the progress shown in the control group, while for MIA Plus the difference in the corresponding progress was 30.9 percent. Likewise, the estimation by instrumental variables shows that the progress relative to the controls, for each tutoring session, was 4.7 and 4 percent measured in both tests, respectively.

[13] The impact results, expressed only in relation to the follow-up score of the controls, without considering the changes between surveys, are shown in Table A10 of the Annex.

Table 7. Change in learning level according to the number of tutoring sessions taken, in terms of progress of the control group

	(1) OLS	(2) 2SLS
IDB-Young (N=2,636)		
In dichotomous variables		
1 to 7 tutoring sessions	**0.032**	
	(0.113)	
8 tutoring sessions	**0.413*** **	
	(0.061)	
Continuous variable	**0.052*** **	**0.047*** **
	(0.008)	(0.015)
MIA PLUS (N=2,636)		
In dichotomous variables		
1 to 7 tutoring sessions	**-0.092**	
	(0.134)	
8 tutoring sessions	**0.309*** **	
	(0.078)	
Continuous variable	**0.038*** **	**.040** **
	(0.010)	(0.018)

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Cost-effectiveness

In addition to the impact found on learning math, the remote tutoring program can be analyzed in terms of its cost and effectiveness. The costs of the program were USD 38 for every 8 tutoring sessions given to each student. The time that the tutors dedicated to the students was 20 minutes per tutoring session, so this cost refers to the total of 3 hours of treatment planned in the program.

This means that for every USD 100 invested per student the gain in math could be as high as 0.69 standard deviations considering that this increase would be proportional to the USD38 invested per student and assuming an estimated increase of 0.033 standard deviations per tutoring session (out of a total of 8). For reference, from the ITT results reported by Carlana (2021) and Melo-Becerra (2021), a gain of 0.48 and 0.32 standard deviations, respectively, for an equivalent investment can be deduced. However, the comparable cost-effectiveness (for those who completed the program) is 17 percent lower than that recorded in Botswana by Angrist et al. (2022), who find an effectiveness of 0.89 standard deviations per USD 100 of investment.

5. Conclusions

The Covid-19 pandemic has had negative consequences for millions of children who interrupted their face-to-face classes and continued their education in emerging distance learning modalities, including learning losses. Various interventions have emerged to mitigate the educational effects of the pandemic, including tutoring programs, some of which have proven effective in pilot schemes in different countries.

This study evaluated a remote tutoring program, planned as a supplement to formal education, using low-tech means such as text messages and phone calls to accelerate the learning of children aged 9 to 14 years in El Salvador and found significant positive impacts. The impact of the intention-to-treat was 0.12 standard deviations on math learning. On the other hand, when considering only the effect on children who took at least one tutoring session, the impact was 0.23 standard deviations, equivalent to an acceleration of 33.2 percent in learning compared to the evolution observed in the control group.

The results suggest that the program has a dose-response effect, where increased learning is strongly associated with the number of tutoring sessions taken. If we include a continuous variable as a treatment, average learning is estimated to be higher by 0.033 standard deviations per tutoring session, and by 0.29 standard deviations when including an indicator variable for completing the program as planned (8 tutoring sessions).

These impacts are of a similar magnitude to those found in other programs, but because these remote tutoring sessions are low-tech, the costs are lower. In this case, the final cost of the program was USD 38 for every 8 tutoring sessions, which is equivalent to an increase of 0.69 standard deviations for every USD 100 of investment, making it a cost-effective intervention to contribute to the recovery of lost learning in case of disruption of face-to-face classes, as in the case of what happened in the wake of the Covid-19 pandemic.

The theory of change was based on the articulation of three dimensions: development of the tutoring relationship, teaching at the appropriate level, and the use of low-tech devices that have high coverage in the region. In this sense, the present study shows that 1) the results obtained are consistent with the available literature regarding the effectiveness of tutoring in increasing fundamental learning (Moeyaert et. al. 2021). The generation of trusting, affective and meaningful relationships between tutees and tutors, the provision of constant feedback, and the willingness to answer doubts and questions are fundamental aspects of the mentoring relationship (Hevia et. al. 2022). In this case, we believe that this relationship could be achieved in a short time because of the one-on-one nature of the tutoring sessions, as well as the challenge of communicating by telephone, as we will see below. 2) These results are consistent with the literature using the "Teach at the right level" principle, both face-to-face (Banerjee, et. al. 2016) as well as remotely (Angrist, Bergman & Matsheng, 2022). In this project, the tutoring started with an initial diagnosis, which identified the appropriate level, worked on the corresponding operations until the tutee demonstrated mastery of the concepts, and only if this happened did the tutee move on to the next operation. Finally, 3) the use of low-tech communication devices, such as telephone calls and text messages, which are in turn the ones with the highest penetration in the region, both in urban and rural areas (IDB, 2021) had a very significant influence. Unlike internet access, where there is a clear digital gap that excludes the unconnected and disproportionately affects the poorest and the rural population

of the region, the telephone has a very extensive penetration among these sectors. The use of low technology made it possible to reach a population that had less support during the pandemic because they were not connected, generating greater equality and educational inclusion. Furthermore, the telephone represented a methodological and novel challenge for teaching and learning. The focus groups showed that, for tutors, being able to teach using only their voice was one of the most challenging aspects.

In terms of the innovative elements of this study, the following stand out: (1) the generation of measures of learning, through face-to-face interviews, which guarantees the quality of the information both for the design of personalized tutoring sessions and for measuring the changes derived from them; (2) to our knowledge, this is the first intervention with experimental evaluation in Latin America using low-tech strategies during the pandemic, which allows us to verify its feasibility for the region in comparison with results obtained in other regions of the world; and (3) the use of two different math learning measurement instruments to verify the robustness of the results obtained.

References

Acevedo, I., Flores, I., Székely, M., Zoido, P. (2022*). ¿Qué ha sucedido con la educación en América Latina durante la Pandemia?* Nota técnica del Banco Interamericano de Desarrollo #2447. http://dx.doi.org/10.18235/0004175

Alegre-Ansuategui, F. J., Moliner, L., Lorenzo, G., & Maroto, A. (2018). Peer Tutoring and Academic Achievement in Mathematics: A Meta-Analysis. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(1), 337-354. https://doi.org/10.12973/ejmste/79805

AMAI (2018). *Nivel Socio Económico AMAI 2018*. *Nota Metodológica*. Asociación Mexicana de Agencias de Inteligencia de Mercado y Opinión. https://www.amai.org/descargas/Nota-Metodolo%CC%81gico-NSE-2018-v3.pdf

Angrist, N., Bergman, P. & Matsheng, M. (2022). Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour 6,* 941–950. https://doi.org/10.1038/s41562-022-01381-z

Angrist, N., D. K. Evans, D. Filmer, R. Glennerster, F. H. Rogers, & S. Sabarwal. (2020). *How to Improve Education Outcomes Most Efficiently? A Comparison of 150 Interventions Using the New Learning-Adjusted Years of Schooling Metric*. Policy Research Working Papers. The World Bank. https://doi.org/10.1596/1813-9450-9450.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, & M. Walton. (2016). "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of 'Teaching at the Right Level' in India". Working Paper 22746. Washington: National Bureau of Economic Research. https://doi.org/10.3386/w22746.

Carlana, M., y La Ferrara, E. (2021). *Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic*. IZA Discussion Paper No. 14094, Available at SSRN: http://dx.doi.org/10.2139/ssrn.3785058

DYGESTIC (2022). *Indicadores.* Indicadores de pobreza y educación. Dirección General de Estadística y Censos. Ministerio de Economía. http://www.digestyc.gob.sv/index.php/servicios/descarga-de-documentos/section/10-indicadores.html

Dietrichson, J., Bøg, M., Filges, T., & Klint Jørgensen, A.-M. (2017). Academic Interventions for Elementary and Middle School Students With Low Socioeconomic Status: A Systematic Review and Meta-Analysis. *Review of Educational Research, 87*(2), 243–282. https://doi.org/10.3102/0034654316687036

Fundación Pro Educación de El Salvador (2021). *Informe De Investigación: Efectos Socioemocionales de la Pandemia por Covid-19 en Estudiantes de Cuarto Grado a Primer Año de Bachillerato del Sistema Educativo Salvadoreño.* Ministerio de Educación Ciencia y Tecnología, El Salvador.

Gobierno del Estado de Sao Paulo (2021). *O impacto da pandemia na Educação Avaliação amostral da aprendizagem dos estudantes.* https://www.educacao.sp.gov.br/wp-content/uploads/2021/04/Apresenta%C3%A7%C3%A3o-Estudo-Amostral.pdf

Gortazar, L., Hupkau, C. y Roldán, A. (2022). *Online tutoring works: Experimental evidence from a program with vulnerable children.* ESADE Center for Economic Policy, Working paper No

2. https://www.esade.edu/ecpol/wp-content/uploads/2022/03/Working_paper_Menttores.pdf

Hassan, H., Islam, A., Siddique, A. y Wang, L. C. (2021). *Telementoring and homeschooling during school closures: A randomized experiment in rural Bangladesh*. Munich Papers in Political Economy. Working Paper No. 04/2021. https://ideas.repec.org/p/aiw/wpaper/13.html

Hernández-Guzmán, L., Bermúdez-Ornelas, G., Spence, S. H., Montesinos, M. J. G., Martínez-Guerrero, J. I., Villalobos, J. A., & Guajardo, J. G. (2010). Versión en español de la Escala de Ansiedad para Niños de Spence (SCAS). *Revista latinoamericana de Psicología*, 42(1), 13-24

Hevia, F. y Vergara-Lope J.(2016). Evaluaciones educativas realizadas por ciudadanos en México: Validación de la Medición Independiente de Aprendizajes. *Innovación Educativa, 16 (70),* pp. 85-110.

Hevia, F. J., S. Vergara-Lope, A. Velásquez-Durán, & D. Calderón. (2022a). "Estimation of the Fundamental Learning Loss and Learning Poverty Related to COVID-19 Pandemic in Mexico". *International Journal of Educational Development* 88 (enero): 102515. https://doi.org/10.1016/j.ijedudev.2021.102515

Hevia, F., Székely, M., Vinacur, T. y Zoido P. (2022). Tutorías remotas: revisión de la literatura. Documento para difusión no. IDB-DP-00944. Sector Social División de Educación, Banco Interamericano de Desarrollo . http://dx.doi.org/10.18235/0004304

Inter-American Development Bank (IDB) (2021). *Informe anual del Índice de Desarrollo de la Banda Ancha: IDBA 2020: Brecha digital en América Latina y el Caribe*. https://doi.org/10.18235/0003337

Karthik, M., A. Singh, & A. Gainimian. (2017). "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India". 500. BREAD Working Paper. BREAD. https://www.ibread.org/working-paper/bread-working-paper-no-500-january-2017/.

Lee D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* 76: 1071–1102.

León, J., Hevia, F. J., Vergara-Lope, S., Vinacur, T., Zoido, P. (2022). *Validación de instrumentos de medición utilizados en el proyecto "tutorías telefónicas" para recuperar y acelerar los aprendizajes.* Nota técnica. BID.

Leung, Kim Chau. (2019). "An Updated Meta-Analysis on the Effect of Peer Tutoring on Tutors' Achievement." *School Psychology International* 40 (2): 200–214. https://doi.org/10.1177/0143034318808832.

Lichand, G., Dória, C. A., Neto, O. L., & Cossi, J. (2021). The Impacts of Remote Learning in Secondary Education: Evidence from Brazil during the Pandemic. MEP. (2021).

Melo-Becerra, L., Ramos-Forero, J., Rodríguez J., Zárate-Solano, H. (2021). *Efecto de la pandemia sobre el sistema educativo: El caso de Colombia.* Borradores de Economía, Núm. 1179. https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/10225/be_1179.pdf

Moeyaert, M., Klingbeil, D., Rodabaugh, E., and Turan, M. (2021). Three-Level Meta-Analysis of Single-Case Data Regarding the Effects of Peer Tutoring on Academic and Social-Behavioral Outcomes for At-Risk Students and Students With Disabilities. *Remedial and Special Education 42* (2): 94–106. https://doi.org/10.1177/0741932519855079.

Monroy-Gómez-Franco, L., Vélez-Grajales, R., López-Calva, L. (2022). The potential effects of the COVID-19 pandemic on learnings. *International Journal of Educational Development. Volume 91*. https://doi.org/10.1016/j.ijedudev.2022.102581

Nickow, A.J., Oreopoulos, P., & Quan, V. (2020). *The Impressive Effects of Tutoring on PreK-12 Learning: A Systematic Review and Meta-Analysis of the Experimental Evidence.* Working Paper 27476. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w27476.

Radhakrishnan, K., Sabarwal, S., Sharma, U., Cullen, C., Crossley, C,. Letsomo, T., Angrist, N. (2021). *Remote learning: Evidence from Nepal during COVID-19.* The World Bank: Policy Brief. https://documents1.worldbank.org/curated/en/906101626938488506/pdf/Policy-Brief.pdf.

Singh, A., Romero M., Muralidharan K. (2022). *Covid-19 Learning Loss and Recovery: Panel Data Evidence from India.* NBER Working Paper No. 30552. *https://www.nber.org/papers/w30552*

Székely, M., Flores, I., Hevia, F., and Calderón, D. (2022). *Measuring Learning Losses from Delayed Return to School: Evidence from Mexico*. Mexicanos Primero-Centro de Estudios Educativos y Sociales (CEES), México, April, 2022.

UNICEF (2021a). *114 millones de estudiantes ausentes de las aulas de América Latina y el Caribe.* Comunicado de Prensa, 14 de marzo de 2021. https://www.unicef.org/elsalvador/comunicados-prensa/114-millones-de-estudiantes-ausentes-de-las-aulas-de-america-latina-y-caribe

UNICEF (2021b). *Casi 2 de cada 3 niños, niñas y adolescentes siguen fuera de las aulas en América Latina y el Caribe.* Comunicado de Prensa, 14 de marzo de 2021. https://www.unicef.org/lac/comunicados-prensa/casi-2-de-cada-3-ninos-ninas-y-adolescentes-siguen-fuera-de-las-aulas-en-america-latina-y-el-caribe

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf

Annex

Table A1. Relationship between math content of interest and placement in the national curriculum of El Salvador

Content of interest	Placement in the curriculum			Indicator of achievement
	Grade	Stage	Week	
Positional value	1st	1	Week 5	1.5 Recognizes the meaning of zero as the absence of elements.
	1st	1	Week 7	3.1 Reads and write the numbers 0 through 10 in ascending order.
	1st	3	Week 1	2.3 Identifies the numbers from 0 to 20, or part of them, on the number line.
	1st	3	Week 7	1.5 Reads, writes and represents with manipulatives from 21 to 29. 1.6 Reads, writes and represents numbers 31 to 39 with manipulatives. 1.7 Reads, writes and represents numbers from 41 to 99 with manipulatives
	1st	3	Week 8	2.1 Forms the numbers up to 99. 2.2 Decomposes the numbers up to 99. 2.3 Writes the tens and units of a 2-digit number in the place value table from its representation with manipulatives.[14]
	1st	3	Week 9	2.4 Composes or decomposes a 2-digit number into tens and units. [15] 2.7 Composes and decomposes the number 100 into tens. 2.8 Counts, reads and the writes numbers from 0 to 100.
Sum of 2 digits plus 2 digits with transformation.	2nd	1	Week 7	2.1 Adds TU + TU, vertically leading to tens. 2.3 Adds TU + TU vertically leading to hundreds. 2.4 Adds TU + TU, vertically leading to tens and hundreds. 2.6 Solves problems involving addition of numbers up to two digits by carrying.
Subtraction of 2 digits minus 2 digits with transformation.	2nd	2	Week 5	2.1 Subtracts TU - TU = TU vertically borrowing from the tens.
	2nd	3	Week 1	2.2 Subtracts TU - TU = U, vertically borrowing from the tens.
Multiplication of 2 digits by 1 digit.	3rd	2	Week 4	3.1 Multiplies vertically TU × U without carrying.
	3rd	2	Week 5	3.2 Multiplies vertically TU × U leading to tens. 3.3 Multiplies vertically TU × U leading to hundreds.
Division of 2 digits by 1 digit with remainder.	3rd	3	Week 7	2.5 Divides vertically TU ÷ U = U with or without residual.
	4th	3	Week 1	1.1 Performs vertical and horizontal divisions TU ÷ U = U Directly using the multiplication table of the divisor.

14 Explicit indicator related to place value.
15 Explicit indicator related to place value.

Table A2. Poverty and education indicators in El Salvador 2013, by department

	Poverty		Years of Schooling			Illiteracy
	Households in Poverty (%)	Population in Poverty (%)	Years of schooling, pop. 10 years or more	Years of schooling, pop. 10-17 years old	Years of schooling, pop. 15-24 years old	Illiteracy rate, pop. 10-17 years old
Ahuachapán	41.4	46.8	5.8	5.5	8.3	3.0
Santa Ana	32.2	38.3	6.4	5.6	8.4	2.3
Sonsonate	32.3	38.6	6.5	5.6	8.7	2.2
Chalatenango	36.8	41.0	6.2	5.9	8.7	2.3
La Libertad	27.0	31.7	7.4	5.7	9.0	2.5
San Salvador	19.9	23.3	8.7	6.1	10.1	1.1
Cuscatlán	32.7	37.0	6.6	5.9	9.0	1.0
La Paz	33.2	37.8	6.5	6.0	8.7	2.4
Cabañas	44.2	51.8	5.4	5.4	7.7	2.5
San Vicente	35.6	41.2	6.6	6.1	9.0	1.9
Usulután	38.2	42.2	6.3	5.9	8.6	2.6
San Miguel	31.3	36.7	6.5	5.9	8.8	1.6
Morazán	42.6	51.0	5.4	5.7	8.2	2.2
La Unión	31.7	38.1	5.1	5.3	7.7	3.8
National	29.6	34.8	7.1	5.8	9.0	2.0

Notes: Own preparation based on official data (DYGESTIC, 2022), for the most recent year (2013) with publicly available information.

Table A3. Balance of characteristics between children in treatment and control groups, at baseline

	Control	Treatment	Difference
Female	0.524	0.512	0.013
	(0.012)	(0.012)	(0.0170)
Age	11.843	11.782	0.061
	(0.038)	(0.039)	(0.0546)
Age - 9 to 11	0.419	0.440	-0.020
	[0.012]	[0.012]	(0.0169)
Disability	0.056	0.056	0.000
	(0.006)	(0.006)	(0.0079)
Grades 7 to 9	0.291	0.294	-0.003
	(0.011)	(0.011)	(0.0155)
Face-to-face	0.874	0.857	0.017
	(0.008)	(0.008)	(0.0116)
Enjoys studying a lot	0.845	0.837	0.009
	(0.009)	(0.009)	(0.0125)
A great desire to learn	0.920	0.916	0.003
	(0.007)	(0.007)	(0.0094)
Severe overage	0.052	0.045	0.007
	(0.005)	(0.005)	(0.0073)
Parents who did not graduate high school	0.693	0.683	0.010
	(0.011)	(0.011)	(0.0158)
SES Score	113.465	112.730	0.734
	(1.077)	(1.065)	(1.5145)
High SES	0.133	0.124	0.008
	(0.008)	(0.008)	(0.0114)
Medium SES	0.298	0.308	-0.010
	(0.011)	(0.011)	(0.0157)
Low SES	0.570	0.568	0.002
	(0.012)	(0.012)	(0.0169)
Generalized Anxiety	0.328	0.333	-0.005
	(0.011)	(0.011)	(0.0160)
Obsessive-Compulsive Disorder	0.370	0.369	0.001
	(0.012)	(0.012)	(0.0165)
Social Phobia	0.283	0.284	-0.001
	(0.011)	(0.011)	(0.0154)
Separation Anxiety	0.593	0.584	0.009
	(0.012)	(0.012)	(0.0168)
Panic disorder/Agoraphobia	0.430	0.411	0.019
	(0.012)	(0.012)	(0.0168)
Fear of Injury	0.317	0.302	0.015
	(0.011)	(0.011)	(0.0158)
Chalatenango	0.351	0.349	0.001
	(0.012)	(0.011)	(0.0163)
Cabañas	0.401	0.405	-0.005
	(0.012)	(0.012)	(0.0167)
Morazán	0.248	0.245	0.003
	(0.010)	(0.010)	(0.0147)
Correct Answers (%) IDB-Young	0.308	0.294	0.0139
	(0.007)	(0.007)	(0.0103)
Correct Answers (%) MIA Plus	0.349	0.345	0.003
	(0.005)	(0.005)	(0.0077)
Observations	1720	1720	

Notes: Standard errors in parentheses. Levels of significance: * p< 0.10, ** p < 0.05, *** p < 0.01

Table A4. Balance of (baseline) characteristics for the sample that did not participate in the follow-up

	Control	Treatment	Difference
Female	0.522	0.504	0.018
	(0.024)	(0.026)	(0.035)
Age	11.973	12.016	-0.044
	(0.079)	(0.085)	(0.116)
Age - 9 to 11	0.398	0.365	0.033
	(0.023)	(0.025)	(0.034)
Disability	0.071	0.057	0.014
	(0.012)	(0.012)	(0.017)
Grades 7 to 9	0.334	0.330	0.004
	(0.023)	(0.025)	(0.033)
Face-to-face	0.863	0.853	0.010
	(0.016)	(0.019)	(0.025)
Enjoys studying a lot	0.849	0.801	0.048*
	(0.017)	(0.021)	(0.027)
A great desire to learn	0.911	0.929	-0.018
	(0.014)	(0.013)	(0.019)
Severe overage	0.050	0.046	0.004
	(0.010)	(0.011)	(0.015)
Parents who did not graduate high school	0.719	0.698	0.021
	(0.022)	(0.024)	(0.032)
SES Score	115.499	113.060	2.439
	(2.129)	(2.286)	(3.129)
High SES	0.144	0.136	0.008
	(0.017)	(0.018)	(0.025)
Medium SES	0.307	0.297	0.010
	(0.022)	(0.024)	(0.033)
Low SES	0.549	0.567	-0.018
	(0.024)	(0.026)	(0.035)
Generalized Anxiety	0.341	0.360	-0.019
	(0.023)	(0.025)	(0.034)
Obsessive-Compulsive Disorder	0.414	0.390	0.025
	(0.024)	(0.025)	(0.035)
Social Phobia	0.304	0.316	-0.012
	(0.022)	(0.024)	(0.033)
Separation Anxiety	0.645	0.591	0.054
	(0.023)	(0.026)	(0.034)
Panic disorder/Agoraphobia	0.506	0.463	0.043
	(0.024)	(0.026)	(0.035)
Fear of Injury	0.339	0.327	0.012
	(0.023)	(0.025)	(0.033)
Chalatenango	0.558	0.627	-0.068**
	(0.024)	(0.025)	(0.035)
Cabañas	0.254	0.232	0.022
	(0.021)	(0.022)	(0.030)
Morazán	0.185	0.139	0.046*
	(0.019)	(0.018)	(0.026)
Correct Answers (%) Bid-Young	0.287	0.275	0.012
	(0.014)	(0.015)	(0.020)
Correct Answers (%) MIA Plus	0.336	0.334	0.002
	(0.011)	(0.011)	(0.015)
Observations	437	367	

Notes: Standard errors in parentheses. Levels of significance: * p< 0.10, ** p < 0.05, *** p < 0.01

Table A5. Balance of characteristics between children who did and did not answer the follow-up questionnaire

	Loss of sample	In Follow-up	Difference
Female	0.514	0.519	-0.006
	(0.018)	(0.010)	(0.020)
Age	11.993	11.758	0.235***
	(0.058)	(0.031)	(0.064)
Age - 9 to 11	0.383	0.443	-0.060***
	(0.017)	(0.010)	(0.020)
Disability	0.065	0.054	0.011
	(0.009)	(0.004)	(0.009)
Grades 7 to 9	0.332	0.280	0.052***
	(0.017)	(0.009)	(0.018)
Face-to-face	0.858	0.868	-0.010
	(0.012)	(0.007)	(0.014)
Enjoys studying a lot	0.827	0.845	-0.018
	(0.013)	(0.007)	(0.015)
A great desire to learn	0.919	0.918	0.001
	(0.010)	(0.005)	(0.011)
Severe overage	0.049	0.048	0.000
	(0.008)	(0.004)	(0.009)
Parents who did not graduate high school	0.709	0.681	0.028
	(0.016)	(0.009)	(0.019)
SES Score	114.386	112.704	1.681
	(1.558)	(0.866)	(1.789)
High SES	0.141	0.125	0.016
	(0.012)	(0.006)	(0.013)
Medium SES	0.302	0.303	-0.000
	(0.016)	(0.009)	(0.019)
Low SES	0.557	0.572	-0.015
	(0.018)	(0.010)	(0.020)
Generalized Anxiety	0.350	0.324	0.025
	(0.017)	(0.009)	(0.019)
Obsessive-Compulsive Disorder	0.403	0.359	0.044**
	(0.017)	(0.009)	(0.019)
Social Phobia	0.310	0.276	0.034*
	(0.016)	(0.009)	(0.018)
Separation Anxiety	0.621	0.579	0.042**
	(0.017)	(0.010)	(0.020)
Panic disorder/Agoraphobia	0.486	0.401	0.086***
	(0.018)	(0.010)	(0.020)
Fear of Injury	0.333	0.302	0.031*
	(0.017)	(0.009)	(0.019)
Chalatenango	0.590	0.277	0.313***
	(0.017)	(0.009)	(0.018)
Cabañas	0.244	0.451	-0.208***
	(0.015)	(0.010)	(0.019)
Morazán	0.164	0.272	-0.107***
	(0.013)	(0.009)	(0.017)
Correct Answers (%) Bid-Young	0.282	0.307	-0.025**
	(0.010)	(0.006)	(0.012)
Correct Answers (%) MIA Plus	0.335	0.350	-0.015*
	(0.008)	(0.004)	(0.009)
Observations	804	2,636	

Notes: Standard errors in parentheses. Levels of significance: * p< 0.10, ** p < 0.05, *** p < 0.01

Table A6. Regression of the attrition variable, with the intention-to-treat, controls and their interactions

	(1)	(2)
	Coefficient	*Std. Error.*
ITT	-0.124	(0.157)
ITT * Female	-0.00881	(0.0280)
ITT * Age	0.0143	(0.0126)
ITT * Disability	-0.0809	(0.0607)
ITT * Grades 7 to 9	-0.0422	(0.0439)
ITT * Face-to-face	0.0334	(0.0416)
ITT * Enjoys studying a lot	-0.0620	(0.0384)
ITT * SES score	-0.000404	(0.000325)
ITT * Separation Anxiety	-0.0643**	(0.0287)
ITT *Correct Answers (%) IDB-Young	0.0201	(0.0479)
ITT * Chalatenango	0.0623	(0.0386)
ITT * Cabañas	0.0376	(0.0368)
Female	-0.00545	(0.0199)
Age	0.00807	(0.00881)
Disability	0.0212	(0.0429)
Grades 7 to 9	0.0112	(0.0306)
Face-to-face	0.00852	(0.0302)
Enjoys studying a lot	0.0392	(0.0275)
SES Score	1.68e-05	(0.000227)
Separation Anxiety	0.0476**	(0.0203)
Correct Answers (%) IDB-Young	-0.0422	(0.0339)
Chalatenango	0.210***	(0.0274)
Cabañas	-0.0313	(0.0261)
Constant	0.0381	(0.112)
Observations	3.440	
R-squared	0.088	

Notes: Standard errors in parentheses. Levels of significance: *p< 0.10, **p < 0.05, ***p < 0.01. This table shows the results of a regression in which the dependent variable indicates whether the child left the sample and therefore did not answer the follow-up questionnaire, with dependent variables of the intention-to-treat, controls, and interaction between the intention-to-treat and controls.

Table A7. Balance of characteristics (baseline) for the sample assigned to the treatment group, according to their participation in the Program

	Untreated		Treated		Difference (4) - (2)
	(1)	(2)	(3)	(4)	(5)
	N	Average	*N*	Average	Diff
Female	653	0.508	700	0.519	-0.01
Age - 9 to 11	653	0.433	700	0.484	-0.051*
Disability	653	0.057	700	0.056	0.001
Second Cycle	653	0.297	700	0.273	0.024
Face-to-face	653	0.867	700	0.85	0.017
Enjoys studying a lot	653	0.832	700	0.86	-0.028
A great desire to learn	653	0.904	700	0.921	-0.018
Severe Overage	653	0.049	700	0.04	0.009
Parents who did not graduate high school	653	0.698	700	0.66	0.038
SES Score	653	114.452	700	110.951	3.5
High SES	653	0.135	700	0.109	0.026
Medium SES	653	0.297	700	0.323	-0.026
Low SES	653	0.568	700	0.569	0
Generalized Anxiety	653	0.322	700	0.329	-0.007
Obsessive-Compulsive Disorder	653	0.364	700	0.361	0.003
Social Phobia	653	0.274	700	0.277	-0.003
Separation Anxiety	653	0.574	700	0.589	-0.014
Panic disorder/Agoraphobia	653	0.401	700	0.393	0.008
Fear of Injury	653	0.302	700	0.289	0.013
Chalatenango	653	0.303	700	0.247	0.056**
Cabañas	653	0.438	700	0.466	-0.028
Morazán	653	0.259	700	0.287	-0.028
Correct Answers (%) Bid-Young	653	0.294	700	0.304	-0.01
Correct Answers (%) MIA PLUS	653	0.344	700	0.353	-0.009

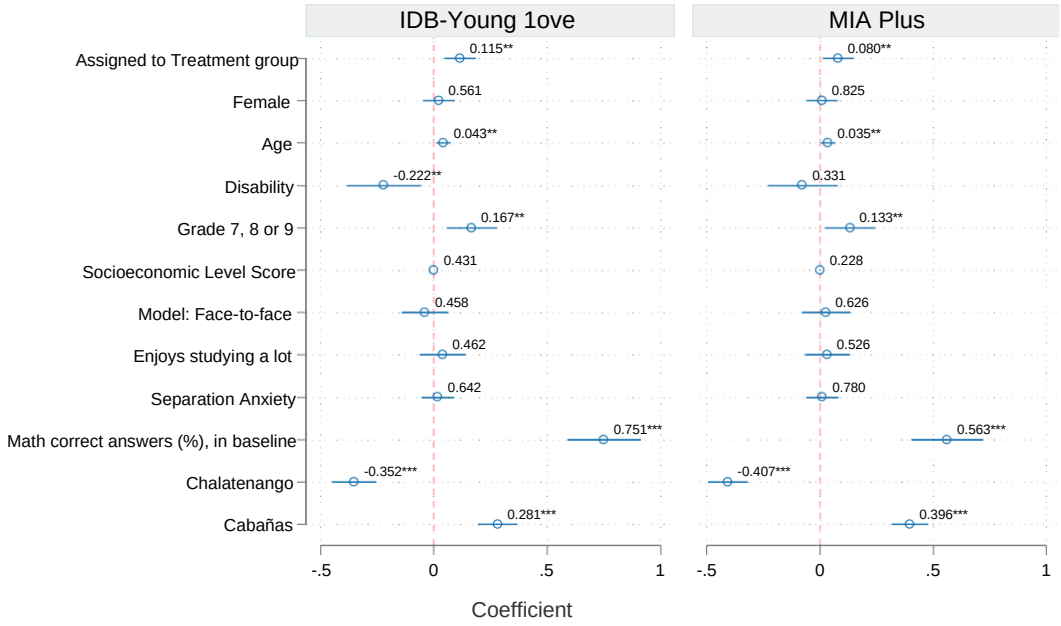
Notes: Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Table A8. Balance of (baseline) characteristics among the children who were treated, by number of tutoring sessions taken

	(1) 1 to 2 tutoring sessions	(2) 3 to 7 tutoring sessions	(3) 8 tutoring sessions	(4) Diff. (1)-(2)	(5) Diff. (1)-(3)	(6) Diff. (2)-(3)
Female	0.423	0.509	0.531	-0.086	-0.109*	-0.023
	(0.059)	(0.067)	(0.021)	(0.089)	(0.063)	(0.069)
Age	11.746	11.877	11.566	-0.131	0.180	0.311
	(0.195)	(0.223)	(0.068)	(0.295)	(0.204)	(0.225)
Age - 9 to 11	0.437	0.386	0.500	0.051	-0.063	-0.114
	(0.059)	(0.065)	(0.021)	(0.088)	(0.063)	(0.069)
Disability	0.127	0.123	0.040	0.004	0.087***	0.083***
	(0.040)	(0.044)	(0.008)	(0.059)	(0.027)	(0.029)
Grades 7 to 9	0.310	0.316	0.264	-0.006	0.046	0.052
	(0.055)	(0.062)	(0.018)	(0.083)	(0.056)	(0.062)
Face-to-face	0.845	0.895	0.846	-0.050	-0.001	0.049
	(0.043)	(0.041)	(0.015)	(0.061)	(0.045)	(0.050)
Enjoys studying a lot	0.859	0.737	0.872	0.122*	-0.013	-0.136***
	(0.042)	(0.059)	(0.014)	(0.070)	(0.042)	(0.048)
A great desire to learn	0.944	0.877	0.923	0.066	0.021	-0.046
	(0.028)	(0.044)	(0.011)	(0.050)	(0.033)	(0.038)
Severe overage	0.042	0.105	0.033	-0.063	0.009	0.072***
	(0.024)	(0.041)	(0.007)	(0.045)	(0.023)	(0.027)
Parents who did not graduate high school	0.648	0.737	0.654	-0.089	-0.006	0.083
	(0.057)	(0.059)	(0.020)	(0.083)	(0.060)	(0.066)
SES Score	111.873	117.193	110.215	-5.320	1.658	6.978
	(5.286)	(6.080)	(1.809)	(8.030)	(5.461)	(6.042)
High SES	0.141	0.123	0.103	0.018	0.038	0.020
	(0.042)	(0.044)	(0.013)	(0.061)	(0.039)	(0.043)
Medium SES	0.296	0.351	0.323	-0.055	-0.028	0.027
	(0.055)	(0.064)	(0.020)	(0.083)	(0.059)	(0.065)
Low SES	0.563	0.526	0.573	0.037	-0.010	-0.047
	(0.059)	(0.067)	(0.021)	(0.089)	(0.062)	(0.069)
Generalized Anxiety	0.423	0.421	0.308	0.001	0.115*	0.113*
	(0.059)	(0.066)	(0.019)	(0.089)	(0.059)	(0.065)
Obsessive-Compulsive Disorder	0.423	0.439	0.346	-0.016	0.076	0.092
	(0.059)	(0.066)	(0.020)	(0.089)	(0.060)	(0.066)
Social Phobia	0.282	0.281	0.276	0.001	0.005	0.004
	(0.054)	(0.060)	(0.019)	(0.081)	(0.056)	(0.062)
Separation Anxiety	0.676	0.544	0.582	0.132	0.094	-0.038
	(0.056)	(0.067)	(0.021)	(0.086)	(0.062)	(0.069)
Panic disorder/Agoraphobia	0.423	0.439	0.385	-0.016	0.038	0.054
	(0.059)	(0.066)	(0.020)	(0.089)	(0.061)	(0.068)
Fear of Injury	0.366	0.211	0.287	0.156*	0.079	-0.076
	(0.058)	(0.054)	(0.019)	(0.081)	(0.057)	(0.062)
Chalatenango	0.296	0.263	0.240	0.033	0.056	0.024
	(0.055)	(0.059)	(0.018)	(0.081)	(0.054)	(0.060)
Cabañas	0.451	0.368	0.477	0.082	-0.027	-0.109
	(0.059)	(0.064)	(0.021)	(0.088)	(0.063)	(0.069)
Morazán	0.254	0.368	0.283	-0.115	-0.030	0.085
	(0.052)	(0.064)	(0.019)	(0.082)	(0.057)	(0.063)
Correct Answers (%) IDB-Young	0.296	0.345	0.301	-0.049	-0.005	0.044
	(0.038)	(0.044)	(0.013)	(0.058)	(0.039)	(0.044)
Correct Answers (%) MIA Plus	0.355	0.368	0.351	-0.013	0.005	0.018
	(0.026)	(0.036)	(0.009)	(0.043)	(0.028)	(0.032)
Observations	71	57	572			

Notes: Standard errors in parentheses. Levels of significance: * p< 0.10, ** p < 0.05, *** p < 0.01

Figure A1. ITT estimation results including control variables



Significance level:* p<0.1, ** p<0.05, *** p<0.01
The magnitudes of the coefficients, omitting the constant, are shown together with the confidence intervals Linear Regression, OLS, Obs 2,636 IDB-Young R2 = 0.144; MIA Plus R2 = 0.153

Table A9. Results (change in probability of correct answers) for effort and fractions questions

	Effort		*Fractions*	
	(1)	(2)	(3)	(4)
	OLS	*2SLS*	*OLS*	*2SLS*
IDB-Young 1ove				
No Controls	**0.012**	**0.022**	**0.007**	**-0.014**
	(0.0193)	(0.0372)	(0.0175)	(0.0339)
With controls	**0.011**	**0.021**	**0.004**	**-0.008**
	(0.0190)	(0.0368)	(0.0168)	(0.0326)
Observations	2,636	2,636	2,636	2,636

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01.

Table A10. Impact results, for variables in levels (test score)

	(1) ITT	(2) TOT	(3) ITT	(4) TOT	(5) Estimate per Session	(6)
	IDB-Young 10ve		**MIA Plus**		**IDB-Young**	**MIA Plus**
Change in Score	**0.223*****	**0.433*****	**0.192****	**0.374****	**.061*****	**.053*****
	(0.067)	(0.128)	(0.086)	(0.166)	(0.018)	(0.023)
Observations	2,636	2,636	2,636	2,636	2,636	2,636
Average number of controls	3.19	3.19	4.5	4.5	3.19	4.5
Percentage Change	7.0%	13.6%	4.3%	8.3%	1.9%	3.4%

Notes: Standard errors in parentheses. Levels of significance: *p < 0.10, **p < 0.05, ***p < 0.01. Score is defined as the total number of correct scores on the test, the maximum number of which is 6 for the IDB-Young 1ove instrument and 9 for MIA Plus. Impacts in columns (1) to (4) are calculated for treatment as a dichotomous variable indicating having taken at least one tutoring session. Columns (5) and (6) show the results of an instrumental variables regression, with the continuous treatment variable indicating number of tutoring sessions, so the results are per tutoring session taken.