ESCUELA DE GOBIERNO

DOCUMENTOS DE TRABAJO 2026/02

Remote tutoring in language: Evidence from Paraguay

Nicolás Acevedo Rebolledo

Gonzalo Almeyda Torres

David Granada Donato

María Lombardi

Victoria Oubiña

Pablo Zoido Lobaton

Marzo 2026

Documentos de trabajo: https://bit.ly/2REorES

UTDT: Av. Figueroa Alcorta 7350, C1428BCW Buenos Aires, Argentina

Remote tutoring in language: Evidence from Paraguay*

Nicolás Acevedo Rebolledo † Gonzalo Almeyda Torres ‡
David Granada Donato § María Lombardi ¶ Victoria Oubiña ‖
Pablo Zoido Lobaton **

March 2026

Abstract

This paper evaluates a randomized remote tutoring program implemented in Paraguay, targeting 1,650 students in grades four through six with low baseline performance in Spanish language. The intervention provided two weekly 30-minute one-on-one tutoring sessions over the phone for eight weeks, using a differentiated instruction model tailored to students' initial diagnostic assessments. Treated students showed significant learning gains: those offered tutoring scored 0.11 standard deviations higher on standardized language tests compared to controls. Effects were consistent across sociodemographic subgroups and baseline achievement levels. Leveraging the random assignment of students to tutors, we estimate individual tutor value added, and find that tutor effects account for 15% of the variation in student outcomes. Tutors in the top quintile have an average value added of 0.38 standard deviations, almost four times the overall effect of the program, underscoring the importance of individual tutor effectiveness in scaling tutoring interventions successfully.

JEL Codes: I20, I24, O15
Keywords: tutoring, education, achievement, skills

*We are grateful for comments from Alex Eble, Felipe Hevia, Juan Leon, Miguel Szekely, and Gisele Teixeira Braun.

†University of Michigan, e-mail: nacev@umich.edu
‡Inter-American Development Bank, e-mail: gonzaloal@iadb.org
§Equilibrium Consulting, e-mail: dgranada@equilibriumbdc.com
¶Universidad Torcuato Di Tella, e-mail: mlombardi@utdt.edu
‖Inter-American Development Bank, e-mail: victoriaou@iadb.org
**Inter-American Development Bank, e-mail: pabloz@iadb.org

1 Introduction

Tutoring is one of the most effective interventions in improving students' academic performance (Fryer, 2017; Nickow et al., 2024; Pellegrini et al., 2021). A likely explanation for its success is tutoring's capacity to focus an instructor's attention on a single student, which allows personalizing instruction tailored to the student's level (Banerjee et al., 2017; Muralidharan et al., 2019; Duflo et al., 2024). However, the effects of tutoring can be diluted when implemented at scale. Kraft et al. (2024), in the largest meta-analysis to date covering 265 interventions, find that scaled programs serving 1,000 or more students achieve an average effect of 0.16 standard deviations on test scores. While this effect size is sizable, it represents a third of the average effect of smaller interventions. This reduction in effectiveness at scale may be due to implementation frictions, such as budget restrictions, decreasing returns when extending the program to students who are less likely to benefit from it, and compromises in key program features—including delivery method, and critically, tutor effectiveness.

Despite the robust evidence on the overall effectiveness of tutoring, a gap in the literature remains to be examined more closely: the role of individual tutors in driving learning outcomes. Understanding which tutors are most effective and why is critical for improving the scalability and cost-effectiveness of tutoring programs. This paper contributes to addressing this gap by examining how tutors influence student learning within a relatively large-scale tutoring intervention.

We evaluate the impact of a randomized remote tutoring pilot implemented in Paraguay, serving 1,650 students in grades four through six who demonstrated low baseline performance. The program delivered two 30-minute one-on-one tutoring sessions per week over the course of eight weeks. All sessions were conducted over the phone, and focused on reinforcing foundational Spanish language skills, in particular fluency and reading comprehension. The program adopted a differentiated instruction model, with tutoring content initially tailored to each student's diagnostic test results. Students advanced to the next level only after mastering the current one, enabling them to reinforce concepts through continued practice until they achieved a solid understanding. The program achieved high takeup: 98% of students in the treatment group attended at least one tutoring session, and on average they participated in 91% of all sessions offered.

We find that the tutoring program significantly improved student learning outcomes. Students offered tutoring scored 0.11 standard deviations higher on standardized language tests than those in the control group. Given the high levels of take-up among the treatment group, we estimate a very similar treatment-on-the-treated effect of 0.12 standard deviations. These effects are robust across subgroups, with no significant heterogeneity detected by baseline performance or sociodemographic characteristics. The program's effectiveness was achieved at relatively low cost, making it highly cost-effective by global standards. Using the equivalent years of schooling (EYOS) framework, we estimate that the intervention generated 0.75 years of learning under business-as-usual conditions and 0.39 EYOS per $100 invested, resulting in a benefit-cost ratio of 34.

Taking advantage of the random assignment of students to tutors, we estimate a model of tutor value added. We find that individual tutors account for 15% of the variation in endline scores for treated students. Furthermore, tutors in the top quintile have an average effectiveness of 0.38 standard deviations, almost four times the overall effect of the program, underscoring the importance of individual tutors in successfully scaling tutoring interventions.

Our paper contributes to the literature studying the effectiveness of tutoring programs. The delivery method for tutoring (in-person vs. remote) has acquired additional relevance since its implementation as a response to learning loss caused by the pandemic. Kraft et al. (2024) suggest that remote tutoring can be equally effective to in-person tutoring, but raise two caveats: the lack of an evaluation directly comparing these delivery methods, and the role of the pandemic in restricting the external validity of the evaluations of remote tutoring interventions carried out in this context. Furthermore, they find that the effects of tutoring on reading and language scores tend to be similar to those of tutoring in math. Among the most notable evaluations of remote tutoring in high-income settings after the pandemic, Ready et al. (2024) find an effect of 0.08-0.26 standard deviations in language scores for an in-school virtual tutoring intervention using the remote platform BookNook, aimed at elementary school students with a duration of 12 weeks and implemented by tutors with previous teaching experience. Carlana and La Ferrara (2024) find an effect of 0.20 standard deviations in math scores for the post-pandemic implementation of the Tutoring Online Program, aimed at middle school students with a duration of 8 weeks and carried out by college students. Additionally, Gortazar et al. (2024) find that the Menttores program had an effect of 0.26 standard deviations for students in middle school, with similar results for two-on-one and one-on-one tutoring.

The effectiveness of remote tutoring holds in low- and middle-income countries, where the interventions are held using SMS or phone calls, as connectivity issues may be expected (Zoido et al., 2024). Several remote tutoring interventions were implemented in this context during the COVID-19 pandemic. Hassan et al. (2024) find an effect of 0.66 standard deviations in language scores for an intervention in Bangladesh, where phone-based tutoring was combined with an intervention aimed at students' mothers. Focusing on numeracy skills, Angrist et al. (2022) evaluate SMS and call-based interventions in Botswana aimed at students between the third and fifth grades during the pandemic, finding an effect of 0.12 standard deviations in math scores. A similar intervention implemented in Kenya, India, Nepal, the Philippines, and Uganda finds average effects of 0.30-0.35 standard deviations (Angrist et al., 2023a). Finally, in several Latin American countries, Albornoz et al. (2025) finds an effect of 0.2 standard deviations in math scores of a phone-based tutoring intervention lasting eight weeks.

This paper provides two main contributions to the literature on tutoring. First, it contributes to the wide and growing body of evidence on the effectiveness of remote tutoring in improving students' academic performance, corroborating its role as a scalable, low-cost policy alternative, even in a post-pandemic context with schools operating normally. Second, it documents the role of individual tutors in driving the effectiveness of tutoring programs. While previous studies have explored mechanisms such as treatment intensity or the tutor-student relationship (White et al., 2023; Carlana and La Ferrara, 2024), this is the first study that explicitly evaluates the contribution of individual tutors to learning gains.

Our findings also contribute to the literature on teacher-value added. Several studies have documented that in preschool and K-12 settings, teachers are among the most important determinants of student learning, with substantial variation in their effectiveness (e.g. Rockoff, 2004; Rivkin et al., 2005; Araujo et al., 2016). We contribute to this literature by showing that tutor effectiveness—like teacher quality—plays a meaningful role in shaping student learning outcomes.

The paper is organized as follows. In Section 2 we describe the program characteristics and experimental design, and in Section 3 we describe the sample and estimation strategy. Section 4 describes the main results, and Section 5 presents our estimates of tutor-value added. Section 6 concludes.

2 Program characteristics and experimental design

The study analyzes the effects of a remote tutoring program in which students received exercises through SMS and interacted remotely with tutors twice a week in 20-30 minute phone calls to improve their Spanish language skills. The intervention was implemented in 57 schools across six departments in Paraguay (Capital, Central, Cordillera, Paraguarí, Misiones and Caaguazú) between March and May 2025.[1] Schools were selected for the

[1] The school year started in February of 2025.

program based on their prior participation in Paraguay's extended school day, a program funded through a loan from the IDB.[2]

Tutoring sessions were aimed at students in the fourth, fifth, and sixth grades, and were designed based on methodological guides aligned with the national curriculum. The pedagogical approach was progressive and student-centered, beginning with basic text parsing and gradually advancing toward the comprehension of more complex literary texts. The tutoring sessions focused on reading fluency, literal and inferential reading comprehension, writing, grammar (particularly sentence structure), punctuation, and vocabulary acquisition. Importantly, the intervention followed a differentiated instruction approach, as students began at the level of the diagnostic test where they first failed, and progressed to more advanced material only once they had mastered the prior content. The phone format was chosen due to its reach and flexibility, especially in zones with limited Internet connectivity.

The program initially administered a contextualized version of the MIA-Plus test to the 5,839 students between fourth and sixth grade that were present on the day of the test in the 57 schools participating in the study.[3] The MIA-Plus test is administered one-on-one, and assesses students across five dimensions: oral reading of two syllables, two words, and two short statements; reading a short literary passage followed by a content-related question; and reading a slightly longer non-literary text followed by a comprehension question (Vergara-Lope et al., 2025).[4] The assessment is administered sequentially, and students who fail to demonstrate proficiency at any given level do not proceed to the following question. Beyond serving as the basis for determining program eligibility, the results also informed the differentiated instruction model by establishing the level at which each tutor began working with a student. The assessment was administered during regular school hours by trained proctors. To ensure a quiet and focused testing environment, schools identified appropriate spaces such as empty classrooms where the assessment could be conducted without distractions.

After the baseline examination, 5,550 students who incorrectly answered at least one question (95%) were identified as eligible for tutoring. Informed consent for participation

[2]Most of these schools (80%) also participated in a pilot math tutoring program implemented by the IDB in 2023 (Albornoz et al., 2025). This may help explain the high levels of take-up in the language tutoring program, as schools were already familiar with a similar intervention. However, only sixth grade students in our sample could have participated in the earlier math pilot, since that program targeted students who were in sixth to eighth grade at the time the language intervention was implemented.

[3]Of the 6,246 students enrolled in these grades, 5,851 were present on the test day, and 5,839 agreed to take the test.

[4]The MIA-Plus test was also implemented in Mexico and Honduras. For further details, see Vergara-Lope et al. (2025).

was requested from the students' guardians, resulting in a total of 3,334 students (60%) who participated in the study. Appendix Table A.2 compares the characteristics of students whose guardians did not consent to participate in the study and those that did. For students who did not consent to participate, we only have information on gender, age, department, and performance in the baseline test. On average, study participation was higher for lower-performing students, who were placed in the two initial performance levels, as well as students in the fourth grade. More specifically, participating students scored 0.19 points lower in the baseline test, compared to a mean of 2.324 for students who did not consent to participate, and were 6.4 percentage points more likely to be in the fourth grade. This pattern aligns with evidence that, in the absence of feedback, parents tend to overestimate their children's academic performance (Bergman, 2021). When presented with additional information, they become more engaged, resulting in gains in student achievement.

The intervention was randomized at the school-by-grade cluster level, with stratification at the school level to ensure that each school had at least one grade assigned to treatment, resulting in cluster randomization across 171 clusters. Randomization resulted in 82 clusters assigned to the control group and 89 clusters assigned to the treatment group. Of the 3,334 students in the sample, 1,650 were assigned to the treatment group and 1,684 to the control group.[5]

The intervention was designed so that students started out at the first level they failed on the diagnostic test. Appendix Figure A.1 shows the distribution of students by starting level. Only 22% started at the first two levels (Syllable and Word), while nearly 42% began at the intermediate level (Statement). Finally, 27% started at the Literary Text level, and about 10% at Reading Comprehension. Starting level is correlated with students' grade level (Appendix Figure A.2), with a higher share of students in upper grades beginning at more advanced levels.

Tutors were recruited via an open call by the Dequení Foundation, and followed highly-structured guidelines developed jointly by the IDB, the Paraguayan Ministry of Education and the Dequení Foundation. The Dequení Foundation is a non-government organization with prior experience implementing tutoring programs in Paraguay.[6] Stu-

[5]Importantly, the rate of student consent for participation in the study was similar across treated and control grades.

[6]Dequení previously led the proof-of-concept and pilot programs in math in 2023 (Albornoz et al., 2025), a scale-up of the math program in 2024–2025, and the proof-of-concept for language in 2024. In turn, Dequení Foundation contracted Equilibrium Social Development to design and carry out the evaluation of the program. Data collection involved the recruitment and training of 77 survey takers, who used mobile devices and the SurveyCTO platform to collect structured survey data. To ensure data integrity, the data collection procedure included a daily supervision and monitoring protocol including real-time feedback.

dents were intended to remain with the same tutor throughout the intervention. Initially, there were 152 tutors, each assigned to 10 students, on average. During the intervention, 10 tutors dropped out (affecting 78 students), and their students were reassigned to either one of the remaining 142 tutors or to two additional tutors who joined later. Appendix Table A.3 summarizes the characteristics of the 142 tutors that remained throughout the intervention. Most of them (89%) were female. Almost half were currently working as teachers, 14% were training to become teachers, and the remainder either had a different occupation (28%) or were studying for a different career (9%). There was substantial variation in teaching experience: 46% had less than two years of experience, 25% had three to six years, and 29% had seven or more years of experience.

Throughout the program, tutors called the students' guardians to confirm the session and sent a SMS with a text exercise two days before the tutoring phone call. On the day of the tutoring session, tutors conducted a 30 minute session addressing the previously sent exercise, and conducting additional activities. Finally, the tutors would contact the students' guardian via SMS after the session, and would ask the student about what they learned during the session, and areas where the student struggled.[7] Students had two tutoring sessions per week, with the same tutor. After each session, tutors would send an SMS to the student's guardian and ask the student to reflect on what they had learned and identify areas where they had struggled. Every three sessions, tutors would conduct and intermediate evaluation to assess whether the student was ready to progress to the next level.[8] The program ended after 15 sessions. Once the program ended, a different version of the MIA-Plus test was administered in person to students in the treatment and control groups, to measure learning gains.

Treatment takeup was high for students included in the study, as shown in Figure 1. Specifically, 98% of the students in the treatment group attended at least one tutoring session, and on average, they participated in 91% of the sessions offered.

Appendix Figure A.3 shows the maximum level reached by students of each starting level who participated in the program at the end of the 8 weeks. A student is considered to have reached a specific level if he/she was able to correctly answer the questions on the previous level during the intermediate evaluations.

[7] For families with access to mobile data, communication was carried out via WhatsApp rather than SMS.

[8] Students covering content from the highest level ("Reading comprehension") were not administered intermediate evaluations to assess progression. Instead, they continued receiving tutoring sessions at this level until the program ended.

3 Sample and estimation

Our initial sample consists of 3,334 students, with 1,650 assigned to the treatment group and 1,684 to the control group. Appendix Table A.1 presents the main descriptive statistics of both groups. The sample consists of students with an average age of approximately 10 years, evenly split by gender. Most students come from bilingual households: approximately half speak both Spanish and Guaraní at home, while almost 40% speak only Spanish, and 9% only Guaraní. In terms of parental education, around 40% of parents or guardians have less than a high school education, a similar share have completed high school, and the remainder have attained a university degree or higher. As expected by randomization, the differences between both groups are small and not statistically significant at conventional levels.

The attrition level at endline was very low, with 3,172 (95%) students taking the endline test. Importantly, the endline response rates are very similar in the treatment and control group (94.5% and 95.7%, respectively), and the difference in response rates between groups is not statistically significant (1.2 percentage point difference, with p-value=0.237).

Appendix Figure A.4 displays the maximum level attained by students in the control group at baseline and endline. There is a marked decline in the share of students at the lowest levels (Word or below), alongside a clear increase in the proportion reaching the Statement level or higher. On average, students advanced by 0.58 levels, though there was considerable variation across individuals. These gains likely reflect the natural progression of learning during the school year, as the intervention overlapped with regular classroom instruction. Students receive at least four hours per week of Spanish language classes, and the gains may also reflect the influence of broader policy efforts.[9]

We estimate intent-to-treat effects using the following equation:

$$Score_{igs} = \beta_0 + \beta_1 Treatment_{gs} + \beta_2 Baseline\ Score_{igs} + \gamma X_{igs} + \delta_g + \nu_s + u_{igs} \tag{1}$$

The dependent variable, $Score_{igs}$, measures the endline score of student i in grade g and school s, standardized with respect to the control group. Our main regressor, $Treatment_{gs}$, is a dummy for whether grade g in school s was assigned to the treatment

[9]The Ministry of Education has been implementing the Ñe'ery program since September of 2023, designed to promote reading, writing, and oral expression across all educational levels. The program provides schools with classroom libraries, teaching materials, and activities such as literary gatherings, fairs, and reading marathons, and also includes teacher training and classroom support to strengthen reading comprehension and improve learning outcomes.

group. To improve the precision of our estimates, we control for students' baseline test scores ($Baseline\ Score_{igs}$), the student's age and gender (X_{igs}), as well as grade and school fixed effects—δ_g and ν_s. Importantly, we are able to include school fixed effects because treatment was stratified at the school level to ensure that each school had at least one treated grade. Standard errors are clustered at the school-by-grade level, the unit of randomization. We also estimate the treatment-on-the-treated using the share of tutoring sessions attended by the student as the treatment variable.

4 Results

Table 1 presents our main results. As shown in column 1, we find that students assigned to treatment score 11.3% of a SD higher, and the effect is statistically significant at the 1% level. Panel B shows treatment-on-the-treated estimates. Students who attended all tutoring sessions experienced a 12.3% SD increase in test scores. This estimate is similar to the intent-to-treat effects, which is expected given the high rate of treatment takeup.

Next, we examine where in the endline test score distribution the observed improvements are coming from. Specifically, we estimate equation (1) using as dependent variables a series of dummies denoting whether the student's highest level achieved on the endline assessment corresponds to Syllable, Word, Statement, Literary Text, or Comprehension. The results are reported in columns 2-7 of Table 1. We find that treatment assignment reduces the likelihood of Statement being the highest level attained by 4.9 percentage points (statistically significant at the 1% level), while increasing the likelihood of reaching the two highest levels: Literary Text by 2.5 percentage points and Comprehension by 4 percentage points. This upward shift in the upper tail of the distribution can be clearly seen in Figure 2.

4.1 Heterogeneous effects

We examine whether the intervention had differential effects across student subgroups by testing for heterogeneous treatment effects based on several baseline characteristics. Specifically, we consider gender, whether the primary languages spoken at home are Spanish and Guaraní or only Guaraní (versus only Spanish), whether the caregiver reported that the student has some difficulty with concentration or memory, whether the caregiver has less than a high school education, and whether the student falls below the sample median in household income. The results are presented in Table 2.[10] With the exception of caregiver education, all of these characteristics are negatively correlated with endline test scores in the control group. Our results indicate that the program was similarly effective across a range of student subgroups, regardless of demographic background. Specifically, we find no statistically significant differences in overall treatment effects by gender, home language, reported difficulties with concentration or memory, or household income. We find suggestive evidence that the effects may be larger for students whose caregiver has completed more than a high school education, although this difference is only significant at the 10% level.

Finally, we analyze whether the program's impact varied according to students' initial skill level. Appendix Figure A.5 presents the results. For students who began at the Syllable or Word level, assignment to treatment increases test scores by 8.8% of a SD. For those starting out at the Statement level, the effect size is 10.9% of a SD, while students who began at the two highest levels—Literary text or Comprehension—experience a 13.2% of a SD increase in endline test scores. Although the effect is statistically significant only for students starting at the Statement level or higher, we cannot reject the null hypothesis that the effects are equal to those of students starting at the Syllable/Word level. Overall, the intervention led to widespread gains in test scores, likely due to its differentiated instruction approach, which allowed tutors to tailor activities to each student's needs.

4.2 Cost-effectiveness

This section provides a cost-effectiveness analysis of the pilot. We measure the program's nominal cost using the existing budget data. Implementing the 8-week pilot had a total cost of $190 per student, with a marginal cost of $137. Fixed costs accounted for 28% of the overall cost, as reported in Appendix Table A.5.

We follow the equivalent years of schooling methodology used in Evans and Yuan (2019). In this approach, effect sizes are converted into equivalent years of schooling (EYOS), translating the effect of the intervention into additional years of schooling under standard conditions. The ITT estimate of 0.11 standard deviations in language scores is equivalent to increasing regular schooling by 0.75 years, under the assumption that a year of regular schooling increases scores by 0.16 standard deviations.[11] This result is

[10] The sample size varies slightly across specifications due to non-response in the baseline demographic survey. Importantly, our main estimates are similar if we restrict our sample to the subgroup with information on each of these characteristics (see Appendix Table A.4).

[11] We use Colombia and Bolivia as benchmarks, as there is no comparable data on the effects of schooling on test scores for Paraguay. (Evans and Yuan, 2019).

comparable to the range observed in structured pedagogy interventions, such as teacher training, teacher support, and the provision of instructional materials, in countries like the Philippines and Kenya, which produce gains of 0.59 to 0.85 EYOS (Evans and Yuan, 2019). The intervention's cost effectiveness is 0.39 EYOS per $100, under the assumption of constant returns.

We then convert these gains in equivalent years of schooling into monetary terms. Since an additional year of schooling in Paraguay increases earnings by 10.4%, annual individual earnings are $4.657 (INE, 2025), and the program is equivalent to an additional 0.75 years of schooling, we estimate that the program increases annual earnings by 7.8%, resulting in an additional $363 per year. These annual gains over an individual's work life sum to a net present value of $6,597.[12] Considering the cost of the pilot, the benefit-cost ratio of the intervention is 34 ($6.597/$190), and the internal rate of return is 186%. While this cost-benefit ratio is favorable, the relatively higher implementation cost leads this program to have a lower cost-benefit ratio than comparable programs in Sub-Saharan Africa and South-East Asia, which face average costs per student of $1-$20 (Evans and Yuan, 2019). However, as fixed costs account for 28% of the cost of the program, the cost-effectiveness of the intervention might improve significantly at scale, especially when integrated with existing administrative infrastructure.

To facilitate cross-country comparisons, we calculate the learning-adjusted years of schooling (LAYS), given that the productivity of school systems varies across countries (Filmer et al., 2020). Using the effect of 0.11 standard deviations, we estimate the LAYS under two scenarios. If the effects last only for the current school year, the 8-week program generates an additional 0.15 years of high-quality schooling per student. If we assume that the effects of the program last for the remaining schooling period, the program generates and additional 0.75 LAYS.[13] Based on these assumptions, the program generates between 0.08 and 0.39 LAYS per $100. While this metric is lower than the average of 3.4 LAYS for interventions in Africa and Asia (Angrist et al., 2023b), it can be attributed to the status of Paraguay as a relatively developed middle income country with higher levels of educational achievement.

[12] We assume a discount rate of 5%, that the individual will work for 40 years, and that the returns to education are constant.

[13] Assuming that sixth-graders have an expected additional 5 years of education ahead, and that a high-quality educational intervention increases scores by 0.8 standard deviations (Angrist et al., 2025). Under this scenario, if the effects of the intervention last for a single year, the intervention will result in 0.14 LAYS (0.11/0.8). If the effects are cumulative, the total effect of the intervention will be 0.6 standard deviations.

5 Tutor value added

We turn our attention to the mechanisms driving the results. The literature suggests that one of the main drivers of the impact of tutoring interventions are individual tutors, given the increased importance of the tutor-student relationship in this setting (Fryer Jr and Howard-Noveck, 2020; White et al., 2023). Thus, we leverage the random assignment of tutors to students in this intervention to estimate a model of tutor value added. For this analysis, we use the 142 tutors that remained in the program until the end.

First, we carry out a Shapley decomposition of test scores in terms of individual student characteristics and school and tutor fixed effects for students in the treatment group. The model has a total R^2 of 0.62, of which 50.71% is explained by baseline scores, 3.77% is explained by age, gender and grade, 30.95% is explained by school fixed effects, and the remainder, 14.56%, is explained by individual tutor fixed effects. This fraction of total R^2 explained by individual tutors is comparable to the 10%-20% range found in earlier studies (Rivkin et al., 2005). The predictive power of individual tutors for academic achievement is remarkable, given the comparatively short timeframe of the intervention.

We test this mechanism formally using a fixed-effects model. As the 141 tutors were randomly assigned to students in alphabetic order, individual tutor fixed effects identify the causal effect of exposure to individual tutors, and can be interpreted as the value added of each tutor. We estimate the following equation:

$$Score_{ijgs} = X_{ijgs} + \gamma_j + \mu_s + e_{ijgs}$$

where the dependent variable, $Score_{ijgs}$, represents the endline score for student i assigned to tutor j in grade j in school s, X_{ijgs} is a set of student-level controls (grade, gender, age, and baseline score), γ_j is a set of tutor fixed effects, and μ_s is a set of school fixed effects.

Figure 3 presents a histogram of the estimated tutor-level value added, γ_j. By definition, the average tutor has an effectiveness equal to the average impact of the program and a value added of zero. The spread of individual tutor value added is considerable. For example, tutors in the top quintile have an average value added of 0.38 standard deviations, almost four times the overall effect of the program. All γ_j are jointly significant at the 1% level and are uncorrelated with students' previous achievement or socio-economic characteristics (see Appendix Figure A.6).

However, these estimates should be interpreted with caution. As each tutor was assigned to 10-12 students, on average, the confidence intervals for individual point estimates are correspondingly wide. As a result, these value added estimates should be only interpreted as describing the role of individual tutor skills in determining the overall impact of the program, and not as a tool to identify and incentivize the most effective tutors. Furthermore, these value added estimates are conditional on the quality of the student-tutor match, and may not necessarily generalize.

5.1 Determinants of tutor value added

As a next step, we attempt to find whether there are any characteristics that may predict individual tutor value added, to determine whether there are any demographic factors that may be useful for targeting the recruitment of the most effective tutors. This is challenging, as studies on teacher value added generally find that differences in teacher value-added are weakly correlated with readily observable teacher characteristics (Hanushek and Rivkin, 2006; Rockoff et al., 2011).

We compiled data on tutors' background, such as gender, years of teaching experience, and education levels, as well as a preliminary suitability score compiled during the recruitment stage. We correlate the previously estimated 142 individual tutor fixed effects with their respective demographic characteristics:

$$\hat{\gamma}_j = X_j + e_j$$

As Appendix Figure A.7 shows, neither the suitability score or the years of teaching experience are significant predictors of tutors' effectiveness. We do find suggestive evidence that female teachers have a smaller value added than males (0.11 standard deviations), and that college students have a higher value added than tutors with a university or postgraduate degree (0.09 and 0.10 standard deviations). These differences are only marginally significant (at the 10% level), and should be interpreted with caution, given the limited sample size. These results nonetheless highlight interesting directions that merit further exploration in future research, particularly regarding the extent to which observable characteristics can or cannot be used to identify effective tutors.

6 Conclusions

Remote tutoring is gaining traction as a promising strategy not only for addressing pandemic-related learning loss, but also as a cost-effective complement to regular instruction in low- and middle-income countries. Its flexibility, scalability, and potential for personalization make it an attractive option for education systems seeking to accelerate foundational learning. Our study contributes to this growing evidence base by evaluating an eight-week remote tutoring, phone based, intervention in Paraguay, which delivered two 30-minute one-on-one sessions per week to students in grades four through six. Using a randomized controlled trial design, we find that the program improved student language scores by 0.11 standard deviations, with no evidence of differential effects by gender or household socioeconomic status.

These learning gains were achieved at relatively low cost. Using standard benchmarks, we estimate that the intervention generated 0.75 equivalent years of schooling (EYOS) and up to 0.75 learning-adjusted years of schooling (LAYS), depending on the persistence of effects. The program yielded 0.39 EYOS per $100 and a benefit-cost ratio of 34, suggesting that remote tutoring can deliver substantial returns even in short time frames and resource-constrained contexts. A key contribution of this study is its focus on understanding the mechanisms behind program effectiveness. We show that individual tutors plays a central role: tutor value added explain 15% of the variation in student outcomes.

References

Albornoz, Facundo, Gonzalo Almeyda Torres, María Lombardi, Victoria Oubiña, and Pablo Zoido (2025) "Remote Tutoring in Latin America," CEPR Discussion Paper No. 20272.

Angrist, Noam, Micheal Ainomugisha, Sai Pramod Bathena et al. (2023a) "Building resilient education systems: Evidence from large-scale randomized trials in five countries," NBER Working Paper 31208.

Angrist, Noam, Peter Bergman, and Moitshepi Matsheng (2022) "Experimental evidence on learning using low-tech when school is out," *Nature human behaviour*, 6 (7), 941–950.

Angrist, Noam, Colin Crossley, and Claire Cullen (2023b) "Africa and Asia Successful Global Tutoring Experiences: ConnectEd."

Angrist, Noam, David K Evans, Deon Filmer, Rachel Glennerster, Halsey Rogers, and Shwetlena Sabarwal (2025) "How to improve education outcomes most efficiently? A review of the evidence using a unified metric," *Journal of Development Economics*, 172, 103382.

Araujo, M Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady (2016) "Teacher quality and learning outcomes in kindergarten," *The Quarterly Journal of Economics*, 131 (3), 1415–1453.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton (2017) "From proof of concept to scalable policies: Challenges and solutions, with an application," *Journal of Economic Perspectives*, 31 (4), 73–102.

Bergman, Peter (2021) "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment," *Journal of Political Economy*, 129 (1), 286–322, 10.1086/711410.

Carlana, Michela and Eliana La Ferrara (2024) "Apart but Connected: Online Tutoring, Cognitive Outcomes, and Soft Skills," Working Paper 32272, National Bureau of Economic Research.

Duflo, Annie, Jessica Kiessel, and Adrienne M Lucas (2024) "Experimental Evidence on Four Policies to Increase Learning at Scale," *The Economic Journal*, 134 (661), 1985–2008.

Evans, David K and Fei Yuan (2019) "Equivalent Years of Schooling: A Metric to Communicate Learning Gains in Concrete Terms. Policy Research Working Paper 8752.," *World Bank*.

Filmer, Deon, Halsey Rogers, Noam Angrist, and Shwetlena Sabarwal (2020) "Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education," *Economics of Education Review*, 77, 101971.

Fryer Jr, Roland G and Meghan Howard-Noveck (2020) "High-dosage tutoring and reading achievement: Evidence from New York City," *Journal of Labor Economics*, 38 (2), 421–452.

Fryer, Roland G (2017) "The production of human capital in developed countries: Evidence from 196 randomized field experiments," in *Handbook of economic field experiments*, 2, 95–322: Elsevier.

Gortazar, Lucas, Claudia Hupkau, and Antonio Roldán-Monés (2024) "Online tutoring works: Experimental evidence from a program with vulnerable children," *Journal of Public Economics*, 232, 105082.

Hanushek, Eric A. and Steven G. Rivkin (2006) "Chapter 18 Teacher Quality," 2 of Handbook of the Economics of Education, 1051–1078: Elsevier, https://doi.org/10.1016/S1574-0692(06)02018-6.

Hassan, Hashibul, Asad Islam, Abu Siddique, and Liang Choon Wang (2024) "Telementoring and homeschooling during school closures: a randomised experiment in rural bangladesh," *The Economic Journal*, 134 (662), 2418–2438.

INE (2025) "Promedio de ingreso mensual (en miles de guaraníes) de la población ocupada de 15 y más años de edad en la ocupación principal por año de la encuesta, según área de residencia, sector económico y años de estudio, 2022 - 2024. Anual," https://www.ine.gov.py/publicacion/5/ingresos.

Kraft, Matthew A, Beth E Schueler, and Grace Falken (2024) "What Impacts Should We Expect from Tutoring at Scale? Exploring Meta-Analytic Generalizability," EdWorking Paper No. 24-1031.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian (2019) "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 109 (4), 1426–1460.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan (2024) "The promise of tutoring for PreK–12 learning: A systematic review and meta-analysis of the experimental evidence," *American Educational Research Journal*, 61 (1), 74–107.

Pellegrini, Marta, Cynthia Lake, Amanda Neitzel, and Robert E Slavin (2021) "Effective programs in elementary mathematics: A meta-analysis," *Aera Open*, 7, 2332858420986211.

Ready, Douglas D, Sierra G McCormick, and Rebecca J Shmoys (2024) "The Effects of In-School Virtual Tutoring on Student Reading Development: Evidence from a Short-Cycle Randomized Controlled Trial. EdWorkingPaper No. 24-942.," *Annenberg Institute for School Reform at Brown University*.

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

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

Rockoff, Jonah E, Brian A Jacob, Thomas J Kane, and Douglas O Staiger (2011) "Can you recognize an effective teacher when you recruit one?" *Education finance and Policy*, 6 (1), 43–74.

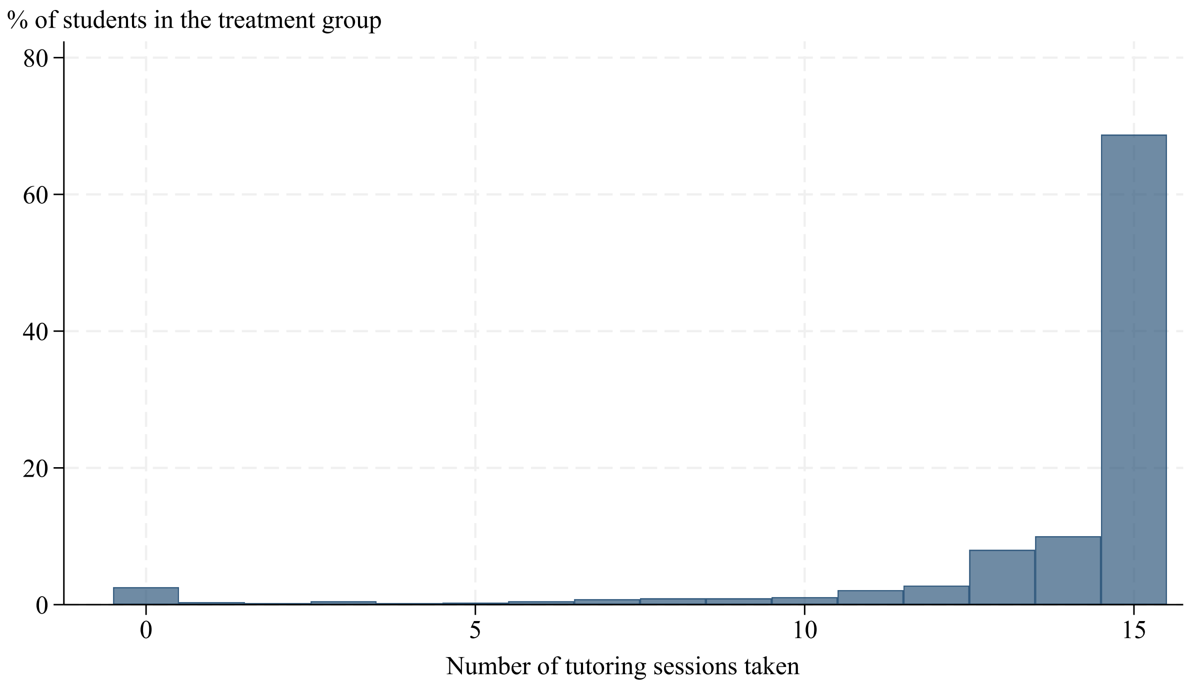
Vergara-Lope, Sergio, Ana Velásquez-Durán, Felipe Hevia, and E. S. González (2025) "Datos de validación del instrumento MIA Plus," working paper, UV – CIESAS.

White, Sara M, Leiah J Groom-Thomas, and Susanna Loeb (2023) "A Systematic Review of Research on Tutoring Implementation: Considerations when Undertaking Complex Instructional Supports for Students. EdWorkingPaper No. 22-652.," *Annenberg Institute for School Reform at Brown University*.

Zoido, Pablo, Iván Flores-Ceceña, Miguel Székely, Felipe J. Hevia, and Eleno Castro (2024) "Remote tutoring with low-tech means to accelerate learning: Evidence for El Salvador," *Economics of Education Review*, 98, 102506.

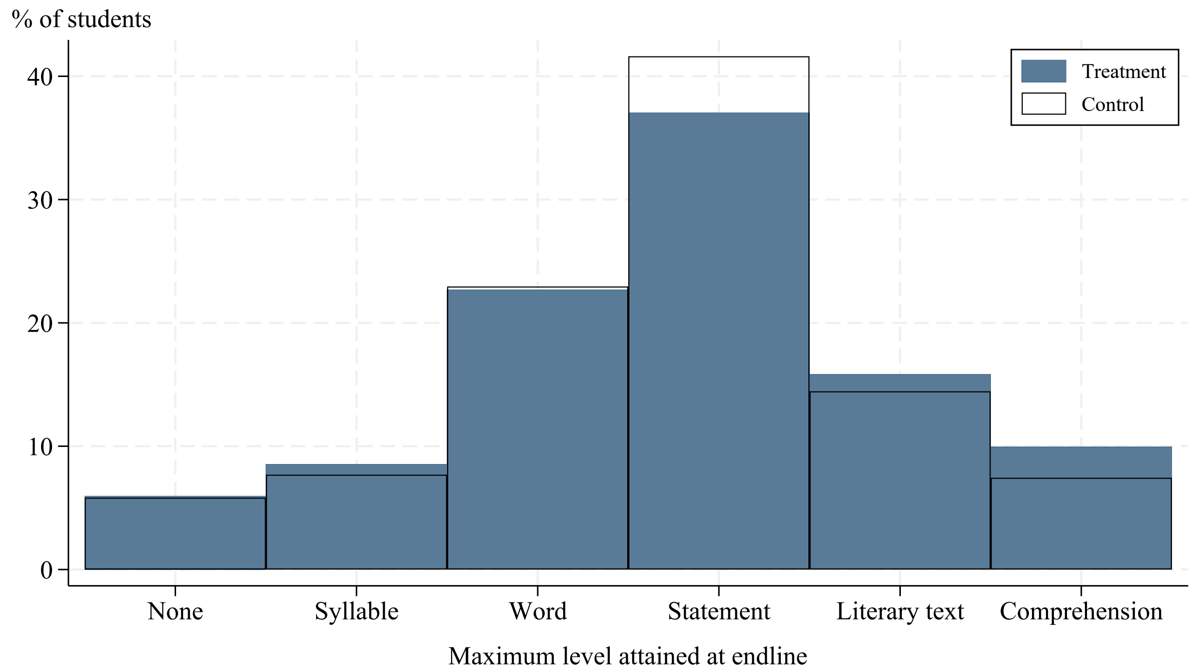
Figures and Tables

Figure 1: Treatment takeup



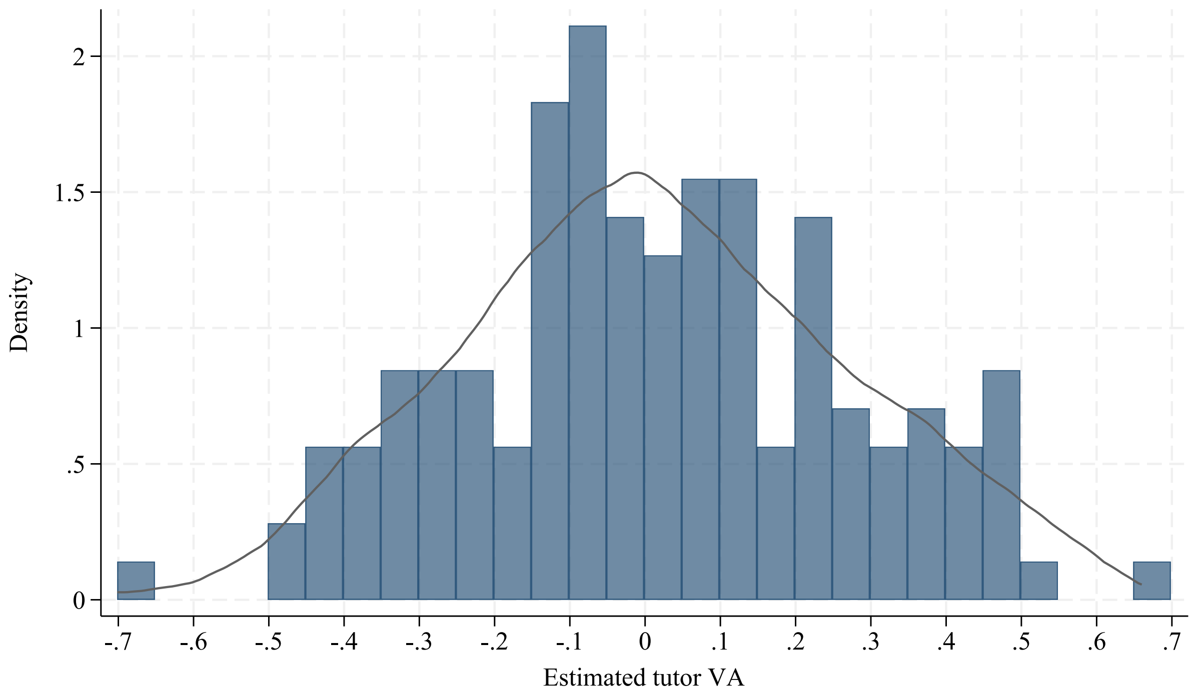
Notes: This figure shows the distribution of the number of tutoring sessions attended by students in the treatment group.

Figure 2: Maximum level attained at endline by the treatment and control groups



Notes: This figure shows the distribution of the maximum level attained by students of the treatment and control group at the endline evaluations.

Figure 3: Distribution of estimates of tutor value added



Notes: This figure shows the distribution of $\hat{\gamma}_j$, the point estimates for tutor fixed effects in the value added model.

Table 1: Impact of tutoring on test scores and maximum level attained

	Total score (z-score)	Maximum level attained: None	Syllable	Word	Statement	Literary text	Comprehension
Panel A: Intent to treat							
Treatment group	0.113***	-0.008	0.001	-0.008	-0.049***	0.025*	0.040***
	(0.036)	(0.007)	(0.008)	(0.013)	(0.018)	(0.014)	(0.012)
Observations	3,172	3,172	3,172	3,172	3,172	3,172	3,172
R^2	0.527	0.239	0.161	0.139	0.119	0.140	0.186
Dep. var. mean (control)	-0.000	0.058	0.077	0.230	0.416	0.145	0.074
Panel B: Treatment on the treated							
Share of tutoring sessions taken	0.123***	-0.009	0.001	-0.008	-0.054***	0.027*	0.043***
	(0.039)	(0.008)	(0.009)	(0.014)	(0.019)	(0.015)	(0.013)
Observations	3,172	3,172	3,172	3,172	3,172	3,172	3,172
R^2	0.433	0.177	0.132	0.027	0.049	0.064	0.064
Dep. var. mean (control)	-0.000	0.058	0.077	0.230	0.416	0.145	0.074
Controlling for baseline scores	✓	✓	✓	✓	✓	✓	✓
Additional controls	✓	✓	✓	✓	✓	✓	✓
School FE	✓	✓	✓	✓	✓	✓	✓

Notes: In Panel A, we present the intent-to-treat estimates, where the main regressor is a dummy for whether the student was assigned to the treatment group. Panel B presents the treatment-on-the treated estimates, with the share of tutoring sessions attended by the student as the main regressor. The dependent variable in column 1 is the student's endline score, standardized with respect to the control group. In the remaining estimations, the dependent variable is a dummy for whether the maximum level attained by the student in the endline test is the level specified in the column header. We control for the student's baseline score, gender, age, grade fixed effects, and school fixed effects. Standard errors clustered at the grade-by-school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2: Intent-to-treat effects – heterogeneous effects by sociodemographic characteristics

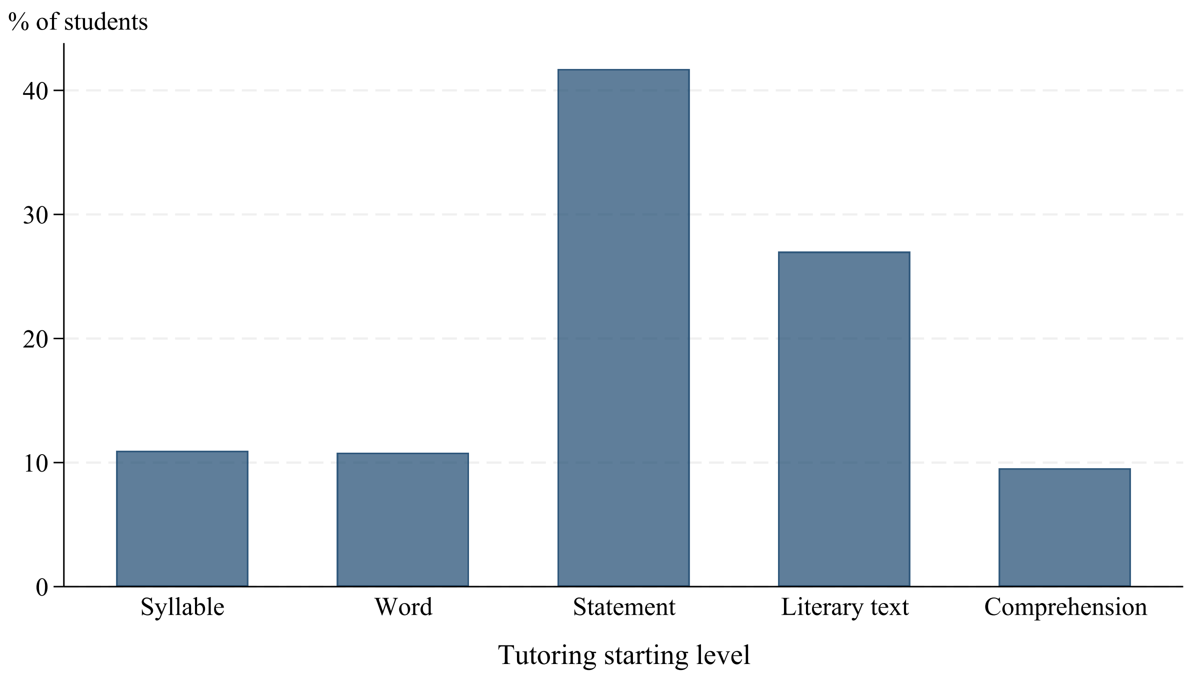
	Characteristic for heterogeneity				
	Male	Main language not only Spanish	Difficulty with concentr./memory	Caretaker educ. < high school	Low income HH
Treatment group	0.082*	0.090*	0.100**	0.146***	0.089**
	(0.042)	(0.046)	(0.044)	(0.043)	(0.038)
Treatment group × Characteristic	0.064	0.016	0.001	-0.109*	-0.019
	(0.053)	(0.062)	(0.051)	(0.060)	(0.053)
Characteristic	-0.091**	-0.068*	-0.111***	-0.052	-0.091**
	(0.041)	(0.040)	(0.035)	(0.037)	(0.040)
Observations	3,172	2,913	2,907	2,913	2,733
R^2	0.527	0.535	0.537	0.537	0.542
P-value (treatment+interaction=0)	0.002	0.024	0.020	0.446	0.166
Controls	✓	✓	✓	✓	✓
School FE	✓	✓	✓	✓	✓

Notes: This table shows the results of regressions where the dependent variable is the student's endline score, standardized with respect to the control group. Our main regressors are a dummy for whether the student was assigned to the treatment group, the characteristic specified in the column header, and the interaction between the two. We control for the student's baseline score, gender, age, grade fixed effects, and school fixed effects. Standard errors clustered at the grade-by-school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

ONLINE APPENDIX

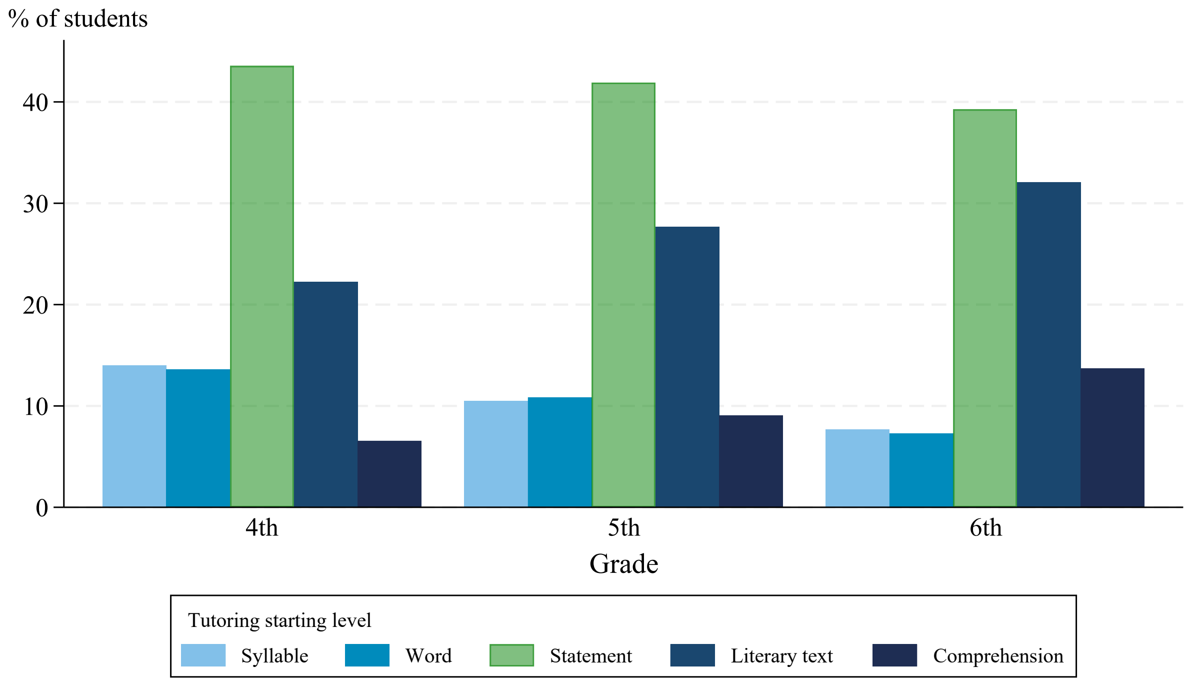
Appendix A Appendix Figures and Tables

Figure A.1: Tutoring starting levels



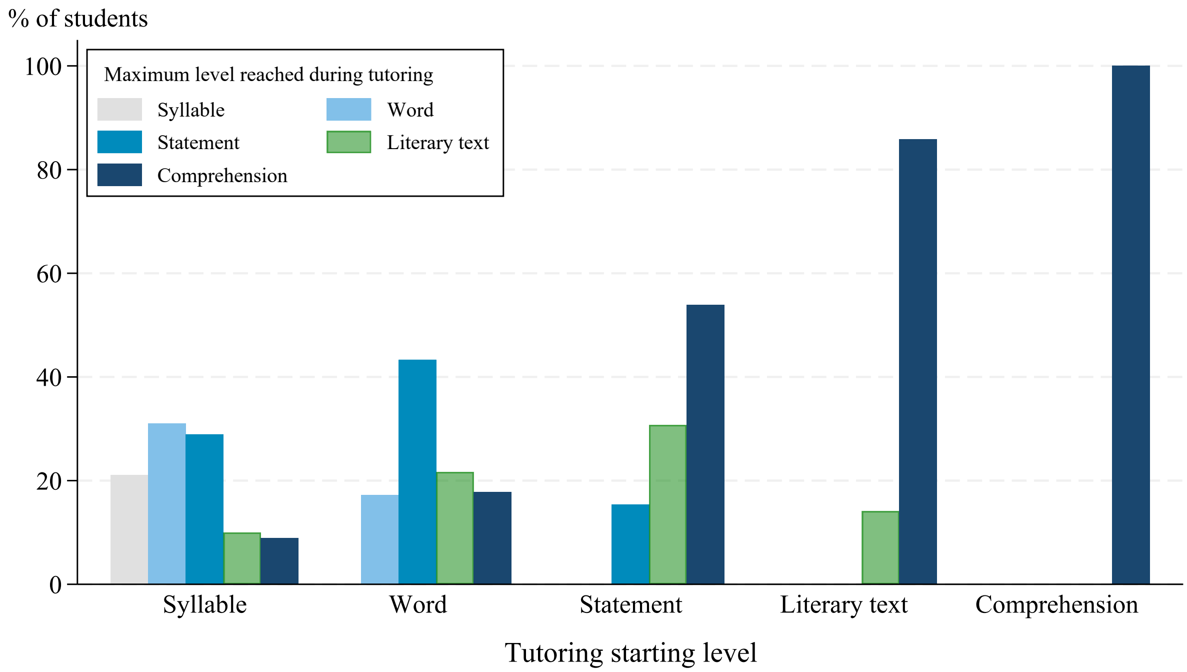
Notes: This figure shows the distribution of students by the level at which they began the tutoring sessions.

Figure A.2: Tutoring starting levels, by grade



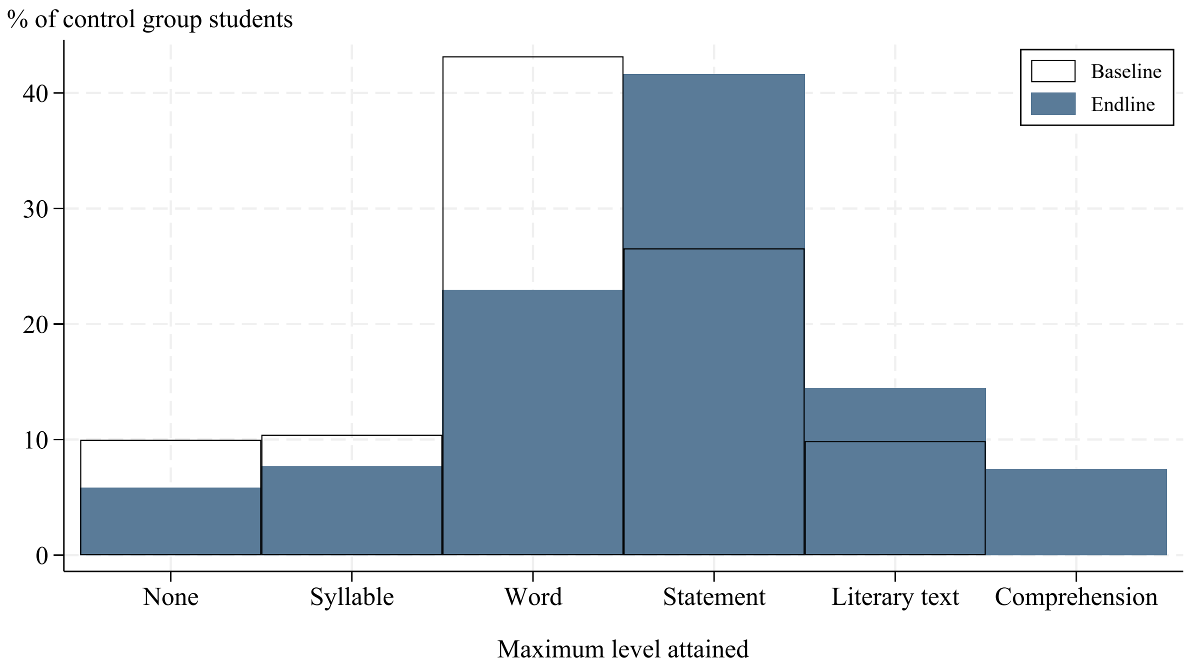
Notes: This figure shows the distribution of students of each grade by the level at which they began the tutoring sessions.

Figure A.3: Maximum level reaches during tutoring, by starting level



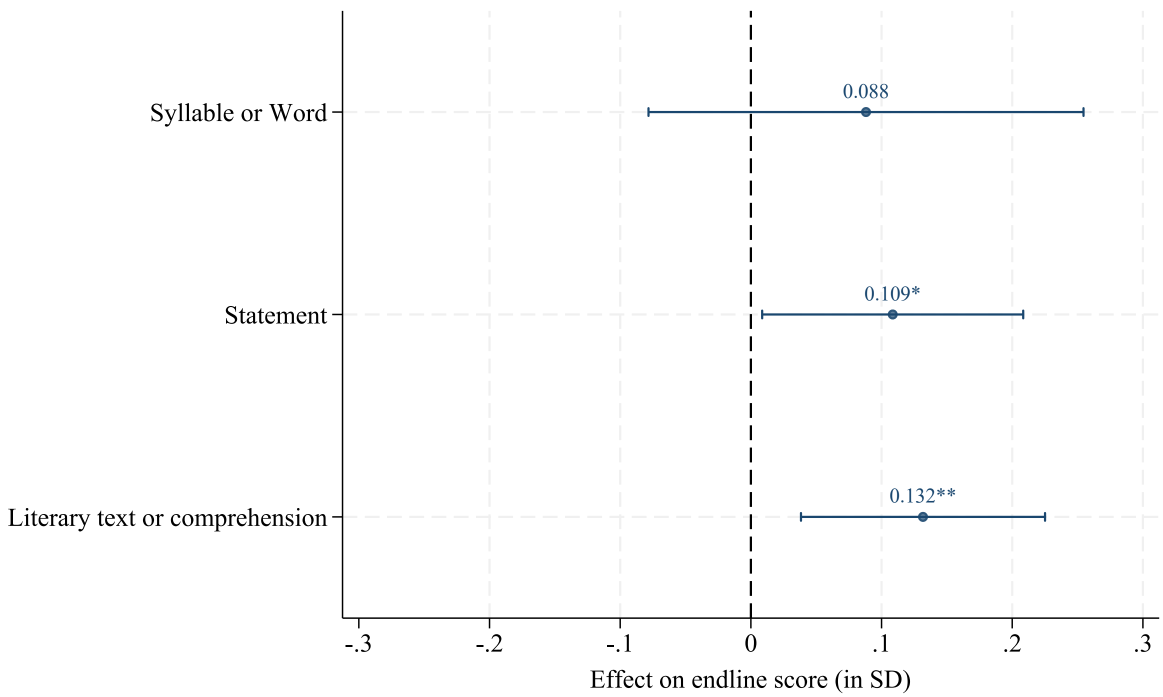
Notes: This figure shows the distribution of the maximum level reached by the students who participated in the tutoring program at the end of the intervention, grouped by their starting level.

Figure A.4: Maximum level attained at baseline and endline by the control group



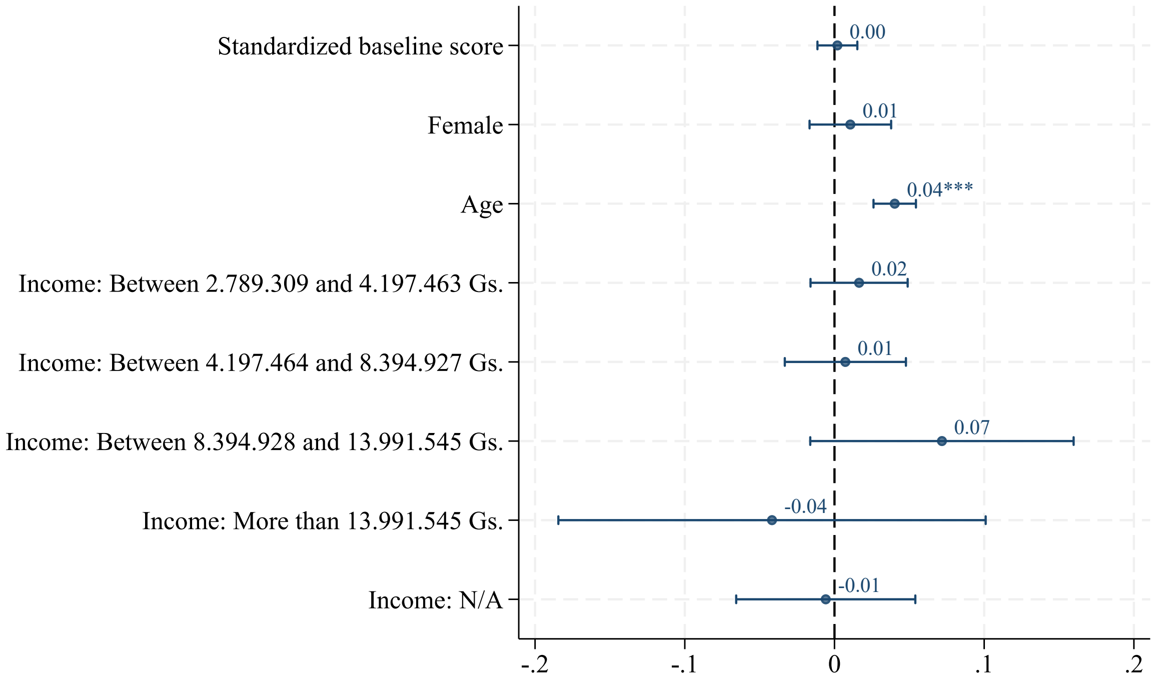
Notes: This figure shows the distribution of the maximum level attained by students of the control group at the baseline and endline evaluations.

Figure A.5: Intent-to-treat effects – heterogeneous effects by tutoring starting levels



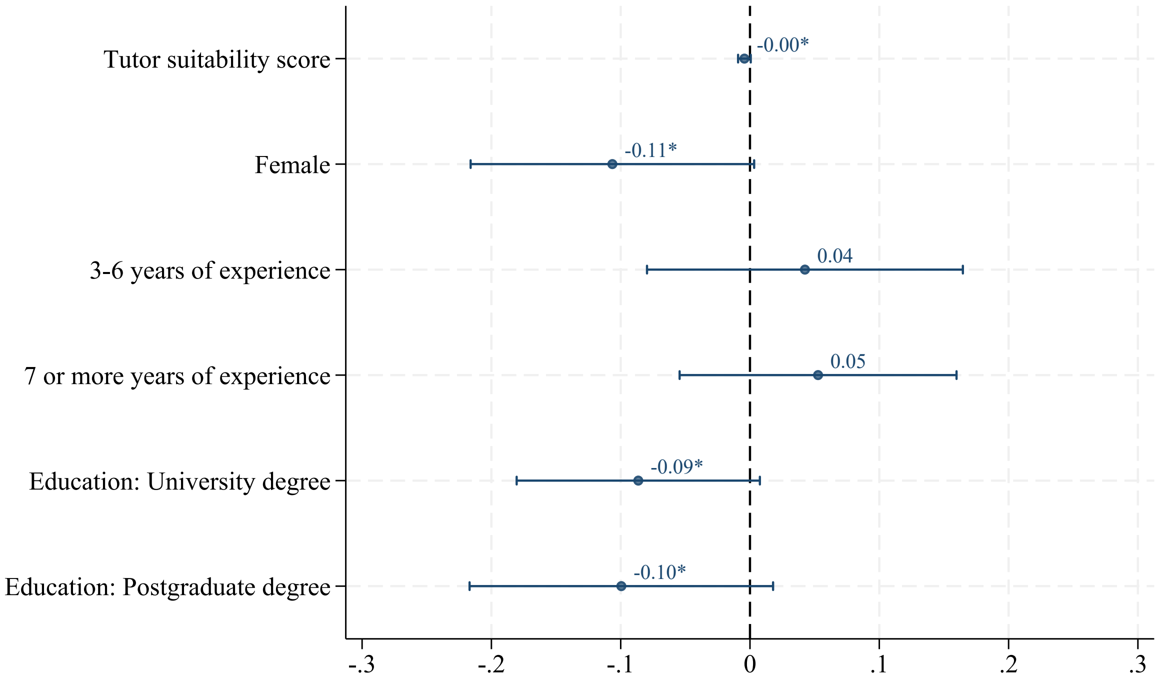
Notes: This table shows the results of a regression where the dependent variable is the student's endline score, standardized with respect to the control group. Our main regressors are dummies for the student started out at Syllable/Word, Statement, or Literary Text/Comprehension, and the interaction of those dummies with a dummy for whether the student was assigned to the treatment group. We report the coefficient and 95%c confidence interval for each of the interactions. We control for the student's baseline score, gender, age, grade fixed effects, and school fixed effects. Standard errors clustered at the grade-by-school level in parentheses.

Figure A.6: Correlation of tutor value added with student characteristics



Notes: This figure shows the correlation of tutor value added point estimates with student baseline characteristics.

Figure A.7: Correlation between tutor value added and tutor characteristics



Notes: This figure shows the coefficients of a regression of the individual tutor fixed effects on their gender, suitability score, teaching experience (the omitted category is less than 3 years of experience), and education (the omitted category is less than a university degree)

Table A.1: Descriptive statistics and balance test

	Control	Treatment	Difference	P-value (diff.)	N
Student characteristics					
Female	0.509	0.491	-0.018	0.331	3,334
Age	9.958	9.809	-0.149	0.298	3,334
Grade: 4th	0.356	0.393	0.037	0.661	3,334
Grade: 5th	0.300	0.342	0.042	0.595	3,334
Grade: 6th	0.344	0.265	-0.079	0.303	3,334
Performance in baseline test					
Correct answers in diagnostic test (0-5)	2.154	2.113	-0.040	0.555	3,334
Tutoring starting level: Syllable	0.101	0.118	0.017	0.274	3,334
Tutoring starting level: Word	0.105	0.112	0.007	0.603	3,334
Tutoring starting level: Statement	0.433	0.401	-0.032	0.212	3,334
Tutoring starting level: Literary text	0.263	0.277	0.014	0.543	3,334
Tutoring starting level: Comprehension	0.099	0.092	-0.006	0.660	3,334
Household and respondent characteristics					
Languages spoken at home: Guaraní and Spanish	0.517	0.490	-0.027	0.340	3,060
Languages spoken at home: Spanish	0.397	0.416	0.019	0.581	3,060
Languages spoken at home: Guaraní	0.085	0.093	0.008	0.666	3,060
Maximum educational attainment: < completed high school	0.401	0.407	0.006	0.847	3,060
Maximum educational attainment: completed high school	0.428	0.405	-0.024	0.288	3,060
Maximum educational attainment: university or more	0.171	0.188	0.018	0.508	3,060
Currently employed	0.639	0.643	0.003	0.888	3,060
School location					
School department: Capital	0.049	0.045	-0.003	0.924	3,334
School department: Cordillera	0.190	0.158	-0.032	0.601	3,334
School department: Caaguazú	0.105	0.070	-0.035	0.538	3,334
School department: Misiones	0.127	0.107	-0.020	0.748	3,334
School department: Paraguarí	0.145	0.169	0.024	0.698	3,334
School department: Central	0.384	0.451	0.067	0.422	3,334
P-value (joint significance)				0.514	

Notes: This table shows the average characteristics at baseline of the control (column 1) and treatment (column 2) groups. Column 3 reports the difference between both groups, and column 4 reports the p-value of the difference in means test. Standard errors clustered at the grade-by-school level.

Table A.2: Descriptive statistics by consent to participate in the study

Variable	No consent	Consent	Difference	P-value (diff.)	N
Student characteristics					
Grade 4	0.310	0.375	0.064***	0.000	6,536
Grade 5	0.348	0.321	-0.028**	0.018	6,536
Grade 6	0.341	0.305	-0.037***	0.002	6,536
Female	0.482	0.500	0.018	0.136	6,536
Performance in baseline test					
Score - Syllable	0.865	0.891	0.026***	0.003	5,839
Score - Word	0.760	0.783	0.023**	0.039	5,839
Score - Statement	0.392	0.365	-0.026**	0.040	5,839
Score - Text	0.193	0.095	-0.097***	0.000	5,839
Score - Understanding	0.115	0.000	-0.115***	0.000	5,839
Baseline score	2.324	2.134	-0.190***	0.000	5,839
School location					
Caaguazú	0.175	0.047	-0.128***	0.000	6,536
Capital	0.089	0.174	0.085***	0.000	6,536
Central	0.082	0.087	0.006	0.402	6,536
Cordillera	0.142	0.117	-0.025***	0.003	6,536
Misiones	0.116	0.157	0.041***	0.000	6,536
Paraguarí	0.396	0.417	0.021*	0.081	6,536

Notes: This table shows the average characteristics at baseline of students that did not consent (column 1) and consented (column 2) to participate in the study. Column 3 reports the difference between both groups, and column 4 reports the p-value of the difference in means test.

Table A.3: Descriptive statistics of tutors

Female	0.887
Occupation: Teacher	0.486
Occupation: Teacher in training	0.141
Occupation: Other	0.282
Occupation: College student	0.092
0-2 years of experience	0.458
3-6 years of experience	0.254
7 or more years of experience	0.289
Education: Current university student	0.296
Education: University degree	0.493
Education: Postgraduate degree	0.211
Number of tutors	142

Notes: This table shows the average characteristics of the tutors that participated in the program.

Table A.4: Intent-to-treat effects – sample with non-missing sociodemographic characteristics

	Sample with non-missing			
	Main language spoken at home	Difficulty with concentr./memory	Caretaker education	HH income
Treatment group	0.102***	0.104***	0.102***	0.087**
	(0.035)	(0.035)	(0.035)	(0.035)
Observations	2,913	2,907	2,913	2,733
R^2	0.535	0.535	0.535	0.541
Controls	✓	✓	✓	✓
School FE	✓	✓	✓	✓

Notes: This table shows the results of regressions where the dependent variable is the student's endline score, standardized with respect to the control group. We limit our sample to the subgroup of students for which we have information on the variable specified in the column header. Our main regressors are a dummy for whether the student was assigned to the treatment group, the characteristic specified in the column header, and the interaction between the two. We control for the student's baseline score, gender, age, grade fixed effects, and school fixed effects. Standard errors clustered at the grade-by-school level in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.5: Cost-benefit analysis

	Pilot
Fixed costs	**$85,256**
Personnel costs	$23,093
Management costs	$62,163
Variable costs	**$220,089**
Tutors and coordinators	$103,177
Technical expenditures (SurveyCTO, devices, mobile data)	$34,305
Outreach to schools and parents	$34,546
Operational costs (materials, transport, etc.)	$48,061
Program size	
Participants	1608
Tutors	144
Students per tutor	11
Schools	57
Total cost	**$305,345**
Variable costs	**$220,089**
Average total cost per student	**$190**
Average variable cost per student	**$137**
Benefits	
Annual wage income in Paraguay	$4657.57
Years in the labor market	40
Program effect (TOT)	0.12
Effect of one year of schooling in language scores (per SD)	0.16
Effect in EYOS	0.75
EYOS per $100	0.39
Return to an additional year of schooling in Paraguay	10.4%
Effect of the program on wage earnings	7.8%
Discount rate	0.05
Annual benefit	$363.29
Net present value of the income stream	$6,597.02
Cost-benefit ratio	34.7
Internal rate of return	186%

Notes: This table summarizes the costs and benefits of the intervention.