# Fostering Early Math Comprehension: Experimental Evidence from Paraguay

Emma Naslund-Hadley
*Inter-American Development Bank*

Susan W. Parker
*Centro de Investigación y Docencia Económicas (CIDE)*

Juan Manuel Hernandez-Agramonte

## Abstract

Research indicates that preschool children need to learn pre-math skills to build a foundation for primary- and secondary-level mathematics. This paper presents the results from the early stages of a pilot mathematics program implemented in Cordillera, Paraguay. In a context of significant gaps in teacher preparation and pedagogy, the program uses interactive audio segments that cover the entire preschool math curriculum. Since Paraguayan classrooms tend to be bilingual, the audio and written materials use a combination of Spanish and Guaraní. Based on an experimental evaluation since the program's implementation, we document positive and significant improvements of 0.16 standard deviations in standardized test scores. The program helped narrow learning gaps between low- and high-performing students, and between students with trained teachers and those whose teachers lack formal training in early childhood education. Moreover, the program improved learning equally among both Guaraní- and Spanish-speaking students. But not all learning gaps narrowed as a result of the program. Although girls improved significantly, boys improved much more, ultimately increasing the gender gap. To close this gender gap, the program has been modified to encourage girls' increased participation in the classroom and general interest in math.

## Keywords

initial education, preschool, early education, mathematics, Paraguay
*JEL:* I21; I28; I29; O15; O31

## Introduction

An increasing body of evidence suggests that the development of pre-math skills at an early age is more important than previously thought in order to foster later mathematical understanding and problem-solving skills. Geary, Hoard, Nugent, & Bailey (2013) find that early knowledge about

**Corresponding Author:**
[6, 7] Emma Naslund-Hadley, Inter-American Development Bank, 1300 New York Avenue, NW, Washington, DC 20577. Email: emman@iadb.org

 **Citation: Naslund Hadley, Emma, Parker, Susan W. & Hernandez-Agramonte, Juan Manuel (2014). Fostering early math comprehension: Experimental evidence from Paraguay. *Global Education Review*, *1* (4). 135-154**

numbers predicts functional numeracy skills in seventh grade. Resnick (1989) argues that developing rudimentary skills in preschoolers, specifically counting abilities, precipitates the retention and automatic use of mathematical concepts in everyday life. In other words, if the use of numbers and counting becomes second nature to children, then these numbers are at their disposal and can be understood in broader contexts (Gersten & Chard, 1999). For example, if a child does not have to ponder whether 9 is greater than 4, but rather knows this to be true, he or she can more easily work with these numbers when learning addition and subtraction. Perhaps even more important, early engagement with numerical concepts is essential for the development of positive attitudes toward math (Clements, Sarama, & Dibiase, 2004), which in turn have been found to be strongly correlated with test scores (House, 2006).

Goldenberg, June, Sword, & Cuoco (2010) have studied the importance of coherence in math across grade levels, showing that each year of math instruction builds on concepts taught in the previous year. Unstructured play alone, others find, is not enough to allow preschool children to reach their full potential in math—they need adult instruction (Clements & Sarama, 2005). Additionally, the literature suggests that rudimentary math skills can be developed at the same time as language and other basic cognitive functions—that is, it is not necessary to focus on language development before promoting the development of math skills (Anuola, Leskinen, Lerkkanen, & Nurmi, 2004).

Studies of preschool math education in the developing world are extremely rare. Those conducted in the United States show that a lack of math education in preschool classrooms does the most harm to children from low socioeconomic backgrounds. That is, among all children who are not exposed to math at the preschool level, those from poorer backgrounds will struggle the most with math in later years (Starkey, Klein, & Wakely, 2004). Studies like these may provide a window into preschool math in the developing world, where children from low socioeconomic backgrounds predominate across school systems.

Inspired by the budding research on early math, the Paraguayan government decided to strengthen its preschool math instruction in 2009. After reviewing a series of math initiatives from across the world, the government chose the Big Math for Little Kids (BMLK) program, which had been implemented in low-income schools in New York. BMLK delivers interactive learning to help young children not only improve their math scores but also to increase their interest in math. BMLK lesson plans and materials were adjusted to Paraguayan conditions and preschool curricular content. The new national preschool math model was named Tikichuela: Mathematics in My School.

It soon became clear that successful implementation of the Tikichuela model required that some of the deficiencies of Paraguayan preschool teachers in our sample, both in knowledge and in pedagogical technique, be addressed. The teachers lacked the skills to translate the lesson plans and activities into the intended classroom activities. Based on the significant positive effects of audio math lessons in Nicaragua,[1] it was decided that Tikichuela would be brought to Paraguayan classrooms through audio lessons, including songs, dances, dramatization of math stories, games, and other interactive activities. These would help reduce the burden on teachers and ensure that all students receive the same instruction, regardless of their teachers' level of pedagogical training and content knowledge.

Additionally, instruction is provided in Spanish and Guaraní, making this a bilingual instruction program. Instruction that is done in a bilingual context has been shown to impact children positively by bridging the learning gap between children that speak two different languages (Wilson et al., 2013). In our program, the audio lessons are provided in Spanish with all key math concepts and central story lines are repeated in Guaraní.

In this paper we report the findings of an experimentally designed evaluation of the Paraguayan pilot during the 2011 academic year, which draws on standardized tests administered in treatment and control groups as well as from surveys of principals, teachers, students, and parents. Section two describes the program background, the instructional approach, and the context in which it was implemented. Section three presents the research design and its implementation, as well as the quality of the data. Section four presents the results and their interpretation. Finally, section five summarizes key insights from the study.

## The Program

### Background

Paraguay is no exception to the generally poor performance of Latin American countries in math and science. A large percentage of those who graduate from high school do not acquire enough knowledge or skills to function well in society. Paraguayan students consistently perform below their peers in other countries in the region. The regional standardized learning test, which assesses the science and math skills of third- and sixth-grade students in 16 countries and territories in Latin America and the Caribbean, reveals that more than half of Paraguayan third-grade students did not attain level II in math. This means that they could not solve simple addition or multiplication problems, extract information from tables, or recognize decimal numbers. In sixth grade a quarter of the students did not reach level II for that grade level, meaning that they were unable to solve problems that required multiplication or division, do addition with fractions, or recognize common geometric shapes (UNESCO-LLECE, 2008).

Results from national assessments confirm the findings of the Segundo Estudio Regional Comparativo y Explicativo (SERCE) test administered in 2006. The National Education Process Evaluation (SNEPE) reveals that, on average, students are falling short of the national goals for math and science proficiency set by the Ministry of Education and Culture (MEC, 2010). Over 50% of third-grade students and some 40% of sixth- and ninth-grade students reach only the most basic level of math achievement.

### The Intervention[2]

Beyond fostering pre-math skills in general, the Tikichuela project was designed to close gaps in learning between students in urban and rural areas, central and peripheral schools within school networks, and multi- and single-grade classrooms. The interactive program was designed to include 108 audio CDs that cover the entire preschool math curriculum. Since Paraguayan classrooms tend to be bilingual, mixing Spanish and Guaraní, the audio programs and written materials are produced using a combination of these languages. Key concepts are repeated in both Spanish and Guaraní. Teachers receive training and in-class tutoring in the interactive audio methodology. As pointed out earlier, the use of bilingual instruction can have positive effects on learning.

In the pilot, the audio lessons were implemented four days a week, with one day set aside to review what had been learned during the week. This extra day gave teachers flexibility to review topics that, according to their observation, the children needed more practice or assistance in addressing. The average duration of each class was 60 minutes, divided into three phases: (a) preparation of the classroom and materials, (b) playing the audio lesson for 30–40 minutes, and (c) additional activities for 15–20 minutes.

The program's introduction was significantly postponed in the first academic year (2011) because of delays in the production of the audio lessons; thus, the planned nine-month implementation time was reduced to five months. That delay, in turn, reduced the number of Tikichuela lessons that could be delivered. Although the program consisted of 108 separate lessons, during the first academic year teachers were able to implement no more than 76 lessons. Although the full program has been implemented since 2011, this paper reports the results from only the first academic year.

### The Context

Baseline tests in March 2011 indicated that preschool children in Cordillera had poor math skills. The average child could name only two out of four geometric shapes, and was unable to recognize four numerals. These deficiencies make it hard for children to succeed in math at the primary level; future learning must build on basic concepts.

Consistent with results from other Latin American countries, Paraguayan students from rural areas and from lower socioeconomic groups were outperformed by students from urban areas and from higher-level socioeconomic households. The baseline data revealed a tendency for the sample to fall into two markedly different groups. The smaller group consisted of urban schools at the center of school networks; these had larger class sizes, single-grade classrooms, and teachers trained in early education. Students in this group obtained scores above the mean across multiple categories. They generally spoke Spanish or both Spanish and Guaraní, came from a family environment with more education, and had previously attended preschool. A second, larger group of rural, peripheral schools was characterized by smaller class sizes, multigrade classrooms, and teachers without adequate training. Students here obtained scores below the mean across multiple categories. They generally spoke Guaraní or both Spanish and Guaraní, came from households with less education, and had not previously attended preschool.

The baseline test indicated a math achievement gap between girls and boys across the entire sample, with boys outperforming girls. Although hardly any Latin American data are available on pre-math skills in very young children, the observed gender gap is in line with math achievement in higher grades across Latin America.[3]

The baseline survey also revealed that preschool teachers felt unprepared to teach math; 94% stated that they had difficulties structuring their math lessons, and 90% said that they were unable to teach all topics in the preschool math curriculum. Additionally, 40% of teachers reported giving math lessons three days or fewer per week, rather than daily as stipulated in the curriculum. These baseline findings suggested that teachers needed support in consistently implementing and completing math lessons.

## Research Strategy: Design, Implementation, and Data

### Design

To be able to estimate whether the project produced the desired effects of raising overall pre-numeracy skills and closing learning gaps, we needed to estimate what would have happened in the absence of the Tikichuela project. To do this, we conducted an experiment in which schools were randomly assigned to either a treatment or a control group. These types of randomized control trials (RCTs) have been used to evaluate the effects of various education inputs, such as textbooks and computers (Barrera-Osorio & Linden, 2009; Malamud & Pop-Eleches, 2011; Cristia, Cueto, Ibarraran, Severin, & Santiago, 2012), scholarships (Glewwe, Hanushek, Humpage, & Ravina, 2011), and tutoring (Banerjee, Cole, Duflo, & Linden, 2007).

The randomized design covered 265 school districts in the department of Cordillera, or approximately 4,500 preschool students and 400 teachers. One hundred and thirty-one schools were randomly selected to receive the treatment, while the remaining 134 schools were designated as the control group. The sample was stratified based on school location (urban-rural), school resources (high-low resources), number of children enrolled, and existence of split sessions schedule (one-two sessions).

To measure the effects, three data collection instruments were used. First, we applied baseline and endline math learning tests at the beginning and the end of the school year. The tests were adapted from the Early Grade Math Assessment (EGMA) developed by the Research Triangle Institute (RTI). To make it possible to assess spillover effects, the tests also included three questions from RTI's Early Grade Reading Assessment (EGRA). The endline test was equivalent to the baseline test, but the level of difficulty was raised to the level expected of a preschool child at the end of the school year. Interviewers administered the tests individually to each preschool child in Spanish or Guaraní, depending on the predominant language of each student. The tests consisted of 14 tasks[4] that took less than 15 minutes to apply to each student. The tests were validated in four schools in the country's central region that were not part of the sample. Second, we surveyed principals, teachers, parents, and students to collect sociodemographic data on the schools and the students' families. Third, we conducted a qualitative evaluation to help us interpret the quantitative findings.

### Data

The baseline information showed that there were no significant differences between the treatment and control groups in terms of school or student characteristics, or the achievement level of the students. But the baseline revealed a significant difference between the two groups in terms of years of teacher experience.

**Table 1. Differences between Treated and Control Groups at Baseline**

| | Treatment (1) | Control (2) | Difference (3) | Observations (4) |
|---|---|---|---|---|
| **Panel A: School Characteristics I** | | | | |
| Number of teachers—preschool | 1.11 | 1.10 | 0.01 | 265 |
| | (0.029) | (0.030) | (0.042) | |
| Preschool enrollment | 16.96 | 17.29 | -0.33 | 265 |
| | (1.408) | (1.393) | (1.981) | |
| Preschool classrooms | 1.22 | 1.25 | -0.03 | 265 |
| | (0.046) | (0.049) | (0.068) | |
| Multigrade preschool classrooms | 1.60 | 1.63 | -0.03 | 265 |
| | (0.043) | (0.042) | (0.060) | |
| Grade-appropriate furniture | 1.19 | 1.22 | -0.03 | 265 |
| | (0.034) | (0.036) | (0.050) | |
| Availability of MOE-provided didactic toys and games | 1.85 | 1.77 | 0.09 | 265 |
| | (0.083) | (0.080) | (0.115) | |
| Ventilation and lighting | 1.09 | 1.11 | -0.02 | 265 |
| | (0.025) | (0.027) | (0.037) | |
| CD player | 1.95 | 1.93 | 0.02 | 265 |
| | (0.051) | (0.053) | (0.073) | |
| Ministry of Education basket of basic preschool materials | 1.05 | 1.02 | 0.02 | 265 |
| | (0.026) | (0.017) | (0.031) | |
| **Panel B: School Characteristics II** | | | | |
| Class size, morning session | 15.20 | 14.50 | 0.70 | 123 |
| | (0.987) | (0.877) | (1.321) | |
| Class size, afternoon session | 13.18 | 13.29 | -0.11 | 201 |
| | (0.698) | (0.650) | (0.954) | |
| Level of formal education of teacher | 4.06 | 4.12 | -0.05 | 289 |
| | (0.051) | (0.061) | (0.079) | |
| Years of experience of teacher | 12.15 | 10.41 | 1.75** | 286 |
| | (0.408) | (0.444) | (0.603) | |
| Language of instruction | 2.53 | 2.53 | 0.01 | 290 |
| | (0.065) | (0.066) | (0.093) | |
| Number of math teacher-training courses | 3.37 | 3.08 | 0.28 | 273 |
| | (0.578) | (0.372) | (0.687) | |
| **Panel C: Performance of Students** | | | | |
| Oral counting | 11.50 | 12.04 | -0.55 | 2907 |
| | (0.242) | (0.224) | (0.330) | |
| Average math score | -0.01 | 0.00 | -0.01 | 2907 |
| | (0.016) | (0.015) | (0.022) | |
| Average literacy score | -0.04 | 0.00 | -0.04 | 2907 |
| | (0.021) | (0.020) | (0.029) | |

*Note:* Estimated standard errors are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Test scores are expressed in standard deviations with respect to the control group.

The attrition rate was low (3.1%). Of the 2,907 students who participated in the baseline test and survey, 2,815 also participated in the final measurement. To ensure that our treatment and control groups were still comparable at the end of the experiment, we used the following regression to test if attrition was systematically related to the treatment:

$$L_{ij} = \delta + \beta \cdot T_j + \varepsilon_{ij} \qquad (1)$$

Where $L_{ij}$ is equal to one if the student *i* in school *j* was not evaluated at the endline, and zero if the student was evaluated. $T_j$ is equal to one if the school *j* was treated, and zero if it is part of the control group. The parameter β will be statistically indistinguishable from zero, if attrition rates were not systematically different between the treatment and control groups. The first column of Table 2 reports the β estimate from equation (1). The estimated coefficient for β is statistically indistinguishable from zero, showing that attrition rates in the treated and control groups are not systematically different.

Although the attrition rate is similar for the treatment and control groups, we must verify if the attrition was orthogonal to the outcomes of interest, that is, if attrition was not correlated to test scores. This would be a problem by implying that even though attrition was statistically the same in treatment and control groups, there is a chance children with lower or higher scores may systematically drop out of the program, biasing the results. We test this empirically through the following regression:

$$Y_{ij} = \delta + \delta_1 \cdot L_{ij} + \beta_1 \cdot T_j + \beta_2 \cdot T_j \cdot L_{ij} + \varepsilon_{ij} \qquad (2)$$

Where $Y_{ij}$ is the standardized test score at baseline for student *i* in school *j*. The other variables are defined as in equation (1). $\beta_1$ estimates the test score difference between the treatment and control groups at baseline if the student did not end up participating in the final measurement, and $\beta_2$ measures the test score difference, also at baseline, between treated and nontreated students that were tested at baseline and at the final measurement. For a causal interpretation of the impact estimations, both coefficients should be statistically indistinguishable from zero. Yet the second column of Table 2 shows that the math results of students for whom we do not have a final measurement are significantly different from those for which we have final measurement.

The negative coefficient indicates that in the case of students who ended up not participating in the endline test, those in the treatment group had lower scores than those in the control group. However, since the attrition rate is low, this result may be random. In fact, this is suggested by the elevated standard errors and by the disappearance of these differences in the more precise measurement presented in the third column. The fourth and fifth columns show that the parameters $\beta_1$ and $\beta_2$ are statistically insignificant for reading comprehension.

During the evaluation, we discovered that one of the 134 control schools had joined the treatment group. As it is not possible to reassign schools, in our analysis the contaminated school is considered a control school, and we report the effects based on the intent to treat.

**Table 2. Attrition Tests for Test Scores**

| Dependent Variables: | Baseline Standardized Scores | | | | |
|---|---|---|---|---|---|
| | Incomplete | Math | | Reading and Comprehension | |
| | (1) | (2) | (3) | (4) | (5) |
| Treated | 0.00 | | | | |
| | (0.01) | | | | |
| Treated x incomplete | | -0.32** | -0.18 | 0.016 | -0.07 |
| | | (0.14) | (0.14) | (0.19) | (0.18) |
| Treated x remaining | | 0.00 | 0.01 | -0.03 | -0.04 |
| | | (0.04) | (0.03) | (0.07) | (0.04) |
| Observations | 2,907 | 2,907 | 2,907 | 2,907 | 2,907 |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Test scores are expressed in standard deviations with respect to the control group.

# Results and Interpretation

## Overall Effects

To assess the program's effect on students' test scores, the following regression was estimated:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \beta \cdot T_j + \varepsilon_{ij,t} \qquad (3)$$

Where $Y_{ij,t}$ is the standardized final measurement for student *i* in school *j*, and $Y_{ij,t-1}$ is the baseline test score in mathematics.[5] The parameter β measures the effect of the program (expressed in terms of standard deviations with respect to the control group) on the test score. The results of the estimated regression are given in Table 3. Column one shows the effects without any controls, column two shows the score when controlling for the baseline score, and column three shows the effects when we control for the baseline score, the strata used in the sample design, and the enumerators. We are particularly interested in column three, since they represent the most robust results.

**Table 3. Overall Results—Mathematics**

| Dependent Variables: | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Overall Effects** | | | |
| Treatment effect | 0.15*** | 0.15*** | 0.16*** |
| | (0.05) | (0.03) | (0.03) |
| Observations | 2,800 | 2,800 | 2,800 |
| **Panel B: Within-School Quartile Effects** | | | |
| Quartile 1 | 0.22*** | 0.20*** | 0.22*** |
| | (0.05) | (0.04) | (0.04) |
| Quartile 2 | 0.11* | 0.12*** | 0.14*** |
| | (0.06) | (0.04) | (0.04) |
| Quartile 3 | 0.16** | 0.18*** | 0.19*** |
| | (0.06) | (0.05) | (0.04) |
| Quartile 4 | 0.09 | 0.08 | 0.09 |
| | (0.08) | (0.06) | (0.05) |
| Observations | 2,800 | 2,800 | 2,800 |
| **Panel C: General Quartile Effects** | | | |
| Quartile 1 | 0.20*** | 0.19*** | 0.20*** |
| | (0.05) | (0.04) | (0.04) |
| Quartile 2 | 0.12*** | 0.12*** | 0.16*** |
| | (0.04) | (0.04) | (0.04) |
| Quartile 3 | 0.21*** | 0.20*** | 0.22*** |
| | (0.05) | (0.05) | (0.04) |
| Quartile 4 | 0.09 | 0.09 | 0.08 |
| | (0.07) | (0.06) | (0.05) |
| Observations | 2,800 | 2,800 | 2,800 |
| **Panel D: Effects by Gender** | | | |
| Boys | 0.22*** | 0.22*** | 0.21*** |
| | (0.06) | (0.04) | (0.04) |
| Girls | 0.08* | 0.10*** | 0.13*** |
| | (0.05) | (0.04) | (0.03) |
| Observations | 2,800 | 2,800 | 2,800 |
| Baseline controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All scores are expressed in standard deviations with respect to the control group.

Students in the pilot program saw an increase in test scores (over those not in the program) of 0.15 standard deviation for column one and two, and 0.16 standard deviation for column three. A 0.16 standard deviation increase means that the students in the treatment group, who on average were in the 50th percentile, moved to the 57th percentile with respect to the distribution of the control group. These findings are most encouraging given the program's short implementation.

One aim of the program was to close achievement gaps among different groups of students. Panel B of Table 4 shows differential effects by quartiles on the baseline exam results. That is, students were ranked according to their baseline test scores within each school, and the distribution was categorized into quartiles. The following regressions were then estimated:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Q1_{ij,t-1} + \delta_3 \cdot Q2_{ij,t-1} + \delta_4 \cdot Q3_{ij,t-1} + \beta_1 \cdot T_j \cdot Q1_{ij,t-1} + \beta_2 \cdot T_j \cdot Q2_{ij,t-1} + \beta_3 \cdot T_j \cdot Q3_{ij,t-1} + \beta_4 \cdot T_j \cdot Q4_{ij,t-1} + \varepsilon_{ij,t} \quad (4)$$

Where $Q_{1ij,t-1}$ is equal to one if the student *i* in school *j* was below the 26th percentile on the baseline test results distribution, and zero otherwise; $Q_{2ij,t-1}$ is equal to one if the student *i* in school *j* was between the 26th and 50th percentile in the baseline measurement, and zero otherwise; $Q_{3ij,t-1}$ is equal to one if the student *i* in school *j* was between the 51st and 75th percentile of the baseline distribution, and zero otherwise; and $Q_{4ij,t-1}$ is equal to one if the student *i* in school *j* was above the 75th percentile, and zero otherwise. The remaining variables are defined as above. The estimates of parameters $\beta_1$, $\beta_2$, $\beta_3$, and $\beta_4$ represent the effects of the program within each quartile of the baseline distribution. Panel B shows that the effects of the program are present in all quartiles except the highest. That is, the program benefits all students with the exception of the highest performers. This means that the program was able to achieve its aim of boosting the performance of those needing the most remedial work in math.

In light of the observed math gender disparities of Latin American students on international standardized tests (the 2009 PISA and 2011 TIMSS), we also analyzed the impacts of the Tikichuela by gender. To obtain the differential effects by gender the following regression was estimated:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Female_{ij} + \beta_1 \cdot T_j \cdot Male_{ij} + \beta_2 \cdot T_j \cdot Female_{ij} + \varepsilon_{ij,t} \quad (5)$$

Where $Female_{ij}$ equals one if student *i* in school *j* is a female and zero otherwise. Similarly $Male_{ij}$ equals one if the student is male, and zero otherwise. Parameters $\beta_1$ and $\beta_2$ capture the additional program effects on boys and girls, respectively. Panel C of Table 3 shows that both boys and girls saw significant increases in their math scores. Notably, Tikichuela girls not only caught up with boys in the control group, but surpassed them despite initially lower scores. But boys across the sample did better than girls, suggesting that the program affected girls differently than it affected boys. From the qualitative data, we know that a combination of cultural factors and attitudes toward gender differences affected the results of the program. Thus the treatment accentuated gender inequality. While a number of possible factors could have contributed to this differential effect, the qualitative evaluation indicated that teachers might have been calling on boys to conduct activities in front of the class more often than girls. Also, there was a widespread perception among the surveyed teachers that boys have more disciplinary problems and are thus in need of more individual attention than girls.

**Impacts by Geographical Location and Type of School**

The sample included 53 urban schools (with 803 students) and 212 rural schools (with 1,997 students), which allowed us to determine whether there are differential impacts by geographical location. We used the following model to estimate these differential effects:

$$Y_{\mathrm{ij,t}} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Urban_{ij} + \beta_1 \cdot T_j \cdot Urban_{ij} + \beta_2 \cdot T_j \cdot Rural_{ij} + \varepsilon_{ij,t} \qquad (6)$$

Where the categorical variable $Urban_{ij}$ equals one if student *i* in school *j* attends a school located in an urban area, and zero otherwise. Similarly, $Rural_{ij}$ equals one if the student is in a school is located in a rural area, and zero otherwise. Parameter $\beta_1$ captures the effects of the program on students in urban areas, while $\beta_2$ captures the additional program effects on students in rural areas with respect to students in urban areas. Table 4 shows the estimated impacts by geographical location.

Table 4 reveals significant effects in both urban and rural areas of 0.12 and 0.18 standard deviations, respectively. Although the effect is larger in rural areas, the difference is not significant. We therefore cannot say that the program contributed to narrowing the achievement gap between students in rural and urban schools.

We were also interested in analyzing if the program helped close the learning gap between central and peripheral schools. To assess these types of effects, we estimated the following model:

$$Y_{\mathrm{ij,t}} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Center_{ij} + \beta_1 \cdot T_j \cdot Center_{ij} + \beta_2 \cdot T_j \cdot Peripheral_{ij} + \varepsilon_{ij,t} \qquad (7)$$

Where $Center_{ij}$ equals one if student *i* in school *j* attends a central school, and zero otherwise. Similarly, $Peripheral_{ij}$ equals one if the student attends a peripheral school, and zero otherwise. Parameter $\beta_1$ captures the effects of the program on students in central schools; while $\beta_2$ captures the additional program effects on students in peripheral schools. Table 5 shows the estimated impacts by school type. Peripheral schools, which typically enjoy fewer resources than those at the center of school networks, saw a significant improvement in scores (0.21 standard deviations). Central schools, on the other hand, did not demonstrate any effect. That is, the Tikichuela program contributed to narrowing the learning gap between the two types of schools.

**Table 4. Results by Geographic Location**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Urban areas | 0.22** | 0.16** | 0.12** |
| | (0.09) | (0.06) | (0.05) |
| Rural areas | 0.12** | 0.14*** | 0.18*** |
| | (0.05) | (0.04) | (0.03) |
| Observations | 2,800 | 2,800 | 2,800 |
| Controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

**Table 5. Results by School Type**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Central school | 0.10 | 0.07 | 0.05 |
| | (0.09) | (0.06) | -0.05 |
| Peripheral school | 0.17*** | 0.18*** | 0.21*** |
| | (0.05) | (0.04) | (0.03) |
| Observations | 2,800 | 2,800 | 2,800 |
| Controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

## Results by Class Size and Grade Composition

Most studies have found that smaller class sizes do not improve overall student performance (Hanushek, 1999), but others have found that class-size reduction programs have positive effects (Kreuger & Whitmore, 2001). In Cordillera the urban schools have an average of 21 students per classroom; in rural areas, the average class size is 14. To assess if the Tikichuela program produced differential effects by class size, we used the following model:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Small_{i\,j,t-1} + \delta_3 \cdot Mediumi_{j,t-1} + \beta_1 \cdot T_j \cdot Small_{i\,j,t-1} + \beta_2 \cdot T_j \cdot Medium_{i\,j,t-1} + \beta_3 \cdot T_j \cdot L\arg e_{ij,t-1} + \varepsilon_{i\,j,t} \qquad (8)$$

Where $Small_{ij}$ equals one if student *i* in school *j* attends a classroom of six or fewer students, and zero otherwise. Similarly, $Medium_{ij}$ equals one if the student attends a classroom of 7 to 16 students, and zero otherwise. Finally, $Large_{ij}$ equals one if student *i* in school *j* attends a classroom of 17 or more students, and zero otherwise. Parameters $\beta_1$, $\beta_2$, and $\beta_3$ capture the program effects on students in small, medium, and large classrooms, respectively. Table 6 reports the effects by class size, showing that class size has a great effect. When controlling for the baseline score, the strata of the sample design, and the enumerators, classes with six or fewer students have an effect of 0.54 standard deviations. We also observe strong effects in classes with 7 to 16 students. But in classes with 17 or more students there was no effect.

As in other Latin American countries, a large proportion of Paraguayan classrooms are multigrade: students in different grades share one teacher. Our sample included 94 schools with multigrade classrooms, 47 of which formed part of the treatment group and 47 of the control group. Since multigrade classrooms present additional instruction challenges and tend to have lower achievement levels than single-grade classrooms, we wanted to assess if the Tikichuela model works in a multigrade setting. To assess if the Tikichuela program produced differential effects by classroom grade composition, we used the following model:

$$Y_{ij,t} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Multigrade_{i\,j} + \beta_1 \cdot T_j \cdot Multigrade_{i\,j} + \beta_2 \cdot T_j \cdot Singlegrade_{i\,j} + \varepsilon_{i\,j,t} \quad \ldots\ldots(9)$$

**Table 6. Results by Class Size**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Class with six or fewer students | 0.48*** | 0.49*** | 0.54*** |
| | (0.15) | (0.13) | (0.11) |
| Class with 7 to 16 students | 0.19*** | 0.20*** | 0.21*** |
| | (0.06) | (0.04) | (0.03) |
| Class with 17 or more students | 0.04 | 0.03 | 0.06 |
| | (0.08) | (0.05) | (0.04) |
| Observations | 2,800 | 2,800 | 2,800 |
| Controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

**Table 7. Results by Classroom Grade Composition**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Multigrade classroom | 0.19** | 0.15** | 0.19*** |
| | (0.08) | (0.07) | (0.05) |
| Single-grade classroom | 0.14** | 0.15*** | 0.16*** |
| | (0.05) | (0.04) | (0.03) |
| Observations | 2,800 | 2,800 | 2,800 |
| Controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Although the multigrade classrooms have larger effect sizes, these differences are not statistically significant. We can therefore conclude that the Tikichuela instructional model works equally well in multi- and single-grade classrooms.

### Literacy Effects

The Tikichuela instructional model does not aim to improve student achievement in other subject areas. But we included tests to assess possible spillover effects on reading and oral comprehension. To assess the program's effect on students' test scores in other subjects, we used equation 3 above, and exchanged the math test scores for literacy scores. The results are described in Table 8.

**Table 8. General Results: Literacy and Oral Comprehension**

| | Naming Letters | Reading | Writing | Oral Comprehension |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A: General Effects** | | | | |
| Effect | -0.05 | 0.05 | 0.07 | 0.11*** |
| | (0.05) | (0.05) | (0.05) | (0.04) |
| **Panel B: Gender Effects** | | | | |
| Boys | -0.01 | 0.05 | 0.09 | 0.11* |
| | (0.07) | (0.06) | (0.06) | (0.06) |
| Girls | -0.08 | 0.05 | 0.03 | 0.10** |
| | (0.06) | (0.06) | (0.05) | (0.05) |
| Observations | 2,800 | 2,800 | 2,800 | 2,800 |
| Controls | Yes | Yes | Yes | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Panel A of Table 8 shows that, with the exception of oral comprehension, the Tikichuela instructional model does not produce any spillover effects. Panel B of the same table shows that the oral comprehension effect is equally distributed between boys and girls. This result is expected since the Tikichuela program focuses strictly on early math skills. The significant spillover into the improvement of oral comprehension is also not surprising since the purpose of the program is to actively engage students in inquiry and communication about math. Sometimes, investment in math can generate spillover effects into other areas. In this case, the model does not seem to improve reading and writing skills. However, the strong focus of the program on listening and communication through audio instruction is reflected in the spillover into increased oral comprehension.

**The Influence of Students' Household Characteristics**

To identify potential differential effects on children from different types of households, parents filled out questionnaires on the language spoken in the home, and the level of education of the parents. Although slightly more parents in the treatment group chose to fill out the questionnaire, the difference was not significant. To assess differential effects by the language spoken at home, we used the following model:

$$Y_{\mathrm{ij,t}} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Guarani_{ij} + \delta_3 \cdot Spanish_{ij} + \beta_1 \cdot T_j \cdot Guarani_{ij} + \beta_2 \cdot T_j \cdot Spanish_{ij} + \beta_3 \cdot T_j \cdot Both_{ij} + \varepsilon_{ij,t} \qquad (10)$$

To assess differential effects by the level of education of the head of household, we used the following model:

$$Y_{\mathrm{ij,t}} = \delta + \delta_1 \cdot Y_{ij,t-1} + \delta_2 \cdot Highschool_{ij} + \beta_1 \cdot T_j \cdot NoHighschool_{ij} + \varepsilon_{ij,t} \qquad (11)$$

Panels A and C of Table 9 report the effects only for those children whose parents filled out the survey questionnaire. Panel A shows that the effect was greater among children living with heads of households without a high school education than among children living with heads of households who have a high school or a higher level of education. The difference in effect size is not significant, however, and we can conclude that the instructional model works equally in both groups of children. Likewise, based on the results reported in Panels B and C, we conclude that the model was successful in both language groups. This is an encouraging as it shows that it is possible to impact learning when two languages are combined in the same lesson.

**Table 9. Effects of the Educational Level and Language of the Student's Household Head**

| Dependent Variables: | Endline Standardized Scores | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Effects—Parental Education** | | | |
| HH head without | 0.17*** | 0.18*** | 0.19*** |
| high school | (0.04) | (0.04) | (0.03) |
| HH head with | 0.12 | 0.12** | 0.12*** |
| high school or higher | (0.08) | (0.06) | (0.05) |
| Observations | 2,448 | 2,448 | 2,448 |
| **Panel B: Effects—Language as Reported by Teacher** | | | |
| Speaks more Guaraní | 0.18*** | 0.15*** | 0.16*** |
| | (0.05) | (0.03) | (0.03) |
| Speaks more Spanish | 0.19*** | 0.16*** | 0.17*** |
| | (0.07) | (0.04) | (0.04) |
| Speaks both equally | 0.07 | 0.14** | 0.18*** |
| | (0.09) | (0.06) | (0.04) |
| Observations | 2,799 | 2,799 | 2,799 |
| **Panel C: Effects—Language as Reported by Head of Household** | | | |
| Speaks more Guaraní | 0.12** | 0.14*** | 0.17*** |
| | (0.06) | (0.05) | (0.04) |
| Speaks more Spanish | 0.19* | 0.15** | 0.14** |
| | (0.10) | (0.07) | (0.06) |
| Speaks both equally | 0.17*** | 0.17*** | 0.19*** |
| | (0.06) | (0.05) | (0.04) |
| Observations | 2,451 | 2,451 | 2,451 |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Test scores expressed in standard deviations with respect to the control group.

## Conclusions

The results of the pilot are very encouraging. An improvement in score of 0.16 standard deviation in less than five months equals or surpasses the effect shown by the majority of studies of interventions to improve academic performance done between 1990 and 2010 (Glewwe et al., 2011). Other programs considered successful in improving childhood math skills achieved similar results, such as the Baisakhi program in India (Krueger and Whitmore, 2001) and the STAR program in the United States (Montsellter, 1995; Folger and Breda 1989), both achieving an increase of 0.18 and 0.15 standard deviations in math scores. Tikichuela's aggregate effect is on par with these programs, especially considering that it was applied for less than one academic year. The positive impact of the program on preschool math skills demonstrates that Tikichuela is more effective in developing math abilities than the traditional methods used by the Cordillera schools.

In light of the many positive findings, it is worth mentioning that our pilot evaluation also helps us understand the limitations of the program and, from those lessons, to improve ours and other similar programs in the future. While the program was successful in the five-month period in which it was implemented, medium- and long-term effects will have to be evaluated.

The Tikichuela methodology is inclusive, benefiting children with initial low math skills (those in the bottom three-quarters on the baseline test) as well as children in low-income schools (peripheral schools) regardless of whether the school has multi- or single-grade classrooms. The program also fosters improvements in children from both language groups (Spanish and Guaraní). Not all learning gaps narrowed as a result of the program, however. The gender gap actually increased, which prompted a modification in the program to encourage girls' increased participation in the classroom and general interest in math. Overall, the results presented here indicate that audio-based instruction can help bridge student learning gaps in a context of large teacher content and pedagogical gaps.

## Notes

1. A 1981 study conducted in Nicaragua provided daily radio mathematics lessons to a group of first-grade classrooms. A second group received mathematics workbooks. After one year, students who had received radio instruction scored 1.5 standard deviations higher than students in a control group, and students given workbooks scored about a third of a standard deviation higher than students in the control group (Heyneman, Jamison, Searle, & Galda, 1981).
2. The program was developed and implemented by the Ministry of Education and Culture (MEC), in collaboration with the Inter-American Development Bank (IDB), the Education Development Center (EDC), the Japan International Cooperation Agency (JICA), and the Organization of Ibero-American States (OEI). The external evaluation was conducted by Innovations for Poverty Action (IPA).
3. In Chile and Honduras—the two countries in the region that participated in the TIMSS 2011 math assessment— boys performed significantly better than girls at both the 4th/6th and 8th/9th grade levels (IEA, 2007). Similarly, a review of the PISA 2009 math assessment (OECD, 2009) showed statistically significant gender differences across all the participating countries in the region (Argentina, Brazil, Chile, Colombia, Mexico, Panama, Peru, and Uruguay).
4. Oral counting, one-to-one correspondence (counting of objects), number identification, missing number recognition, writing numbers, quantity discrimination, successor and predecessor, addition and subtraction, shape recognition, spatial relations, letter identification, reading numbers and words, writing words, oral, and comprehension
5. Test scores were standardized to give the control group a mean of zero and a standard deviation of one.
6. Support for this research was provided by the Inter-American Development Bank.
7. The opinions expressed in this paper 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.

## References


Anuola, K., Leskinen, E., Lerkkanen, M., & Nurmi, J. (2004). Development dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96* (4), 699–713.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2005). Remedying education: Evidence from two randomized experiments in India. NBER Working Paper 11904, National Bureau of Economic Research, Cambridge, MA.

Barrera-Osorio, Felipe & Linden, Leigh L. (2009). "The Use and Misuse of Computers in Education: Evidence from a Randomized Expriment in Colombia" Policy Research Working Paper Series 4836, The World Bank, Washington, DC.

Clements, D. H., & Sarama, J. (2005). Young children and technology: What's appropriate? *67th Yearbook of Technology-supported Mathematics Learning Environments*, 51–73.

Clements, D. H., Sarama, J., & Dibiase, A. (2004). *Engaging Young Children in Mathematics*. London: Psychology Press.

Cristia, J., Cueto, S., Ibarraran, P., Severin, E., & Santiago, A. (2012). Technology and child development: Evidence from the One Laptop per Child Program. IDB Working Paper No. 273, Inter-American Development Bank,

Folger, J., and Breda, C. (1989). Evidence from Project STAR about class size and student achievement. *Peabody Journal of Education*, *67*(1), 17–33.

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *Plos One*, *8* (1), 1–8.

Gersten, R., & Chard, D. 1999. Number sense: rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education*, *33*, 18–28.

Glewwe, P., Hanushek, E., Humpage, S., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010*. Unpublished manuscript, University of Minnesota.

Goldenberg, P., June, M., Sword, S., & Cuoco, A. (2010). Developing mathematical habits of mind. *Mathematics Teaching in the Middle School, 15* (9), 505–509.

Hanushek, E. A. (1999). The evidence on class size. In S. E. Mayer & P. E. Peterson (Eds.), *Earning and learning: How schools matter*. Washington, DC: Brookings Institution Press.

Heyneman, S. P., Jamison, D. T., Searle, B., & Galda, K. 1981. Improving elementary mathematics education in Nicaragua: An experimental study on the impact of textbooks and radio on achievement. *Journal of Educational Psychology*, *73*(4), 556–567.

House, J. D. (2006). Mathematics beliefs and achievement of elementary school students in Japan and the United States: Results from the third international mathematics and science study. *The Journal of Genetic Psychology*, *167*(1), 31–45.

IEA (International Association for Evaluation of Educational Achievement). (2007). Trends in Mathematics and Science Studies (TIMSS). Retrieved from http://timss.bc.edu/TIMSS2007/idb_ug.html.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in early grades on college-test taking and middle school test results: Evidence from Project STAR. *Economic Journal*, *111*(468), 1–28.

Malamud, O., & Pop-Eleches, C. (2011). Home computer use and the development of human capital. *Quarterly Journal of Economics*, *126* (2), 987–1027.

MEC (Ministry of Education and Culture). (2010). Informe preliminar SNEPE. Asunción, Paraguay.

Mosteller, Frederick. 1995. The Tennessee study of class size in the early grades. *The Future of Children*, *5*(2), 113–27.

OECD (Organisation for Economic Co-operation and Development). (2009). Program for International Student Assessment. Retrieved from http://www.oecd.org/edu/pisa/2009.

Patrinos, H. A. and E. Velez (2009). Costs and benefits of bilingual education in Guatemala: A partial analysis. *International Journal of Educational Development*. 29: 594–598.

Resnick, Lauren. 1989. Developing mathematical knowledge. *American Psychologist, 44*(2), 162–169.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19* (1), 99–120.

UNESCO-LLECE (United Nations Educational, Scientific and Cultural Organization–Latin American Laboratory for Assessment of the Quality of Education). (2008). Los Aprendizajes de los Estudiantes de América Latina y el Caribe. First report from SERCE, OREALC/UNESCO, Santiago, Chile.

Wilson, S. J., D. K. Dickinson, and D. W. Rowe (2013). Impact of an Early Reading First program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly*. 28: 578– 592.

**About the Author(s)**

**Emma Naslund-Hadley** is Lead Education Specialist at the Inter-American Development Bank. She manages the Bank's efforts to improve mathematics and natural science education in Latin America and the Caribbean.

**Juan Manuel Hernández-Agramonte** is a Policy Manager for J-PAL Latin America where he is supporting the efforts of regional governments in producing evidence based policy, by building partnerships and providing technical assistance to policymakers for the scale-up and replication of proven effective interventions. Previously he worked closely with researchers and governments in designing and implementing randomized control trials in various sectors.

**Susan W. Parker** is Professor of Economics at the Centro de Investigación y Docencia Económicas (CIDE) in Mexico City and Adjunct Economist at the RAND Corporation. She obtained her doctorate in economics from Yale University. Her research focuses principally on education and health in developing countries and on the design, implementation and evaluation of programs, with a particular emphasis on Latin America.

**Appendix Begins Next Page**

# Appendix

## Annex 1 – Detailed Regression Tables

**Table 3: Overall Results—Mathematics**

| Dependent Variables: | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Overall Effects** | | | |
| Treatment (in standard deviations) | 0.15*** | 0.15*** | 0.16*** |
| | (0.05) | (0.03) | (0.03) |
| Treatment (coefficient in points) | 0.09*** | 0.09*** | 0.09*** |
| | (0.03) | (0.02) | (0.02) |
| Equation y intercept | 0.01 | 0.01 | 0.26 |
| **Goodness of Fit** | | | |
| F(1, 264) | **10.30** | **458.01** | **33.53** |
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.01** | **0.39** | **0.46** |
| Root MSE | 0.66 | 0.52 | 0.49 |
| Observations | 2800 | 2800 | 2800 |
| **Panel C: General Quartile Effects** | | | |
| **Quartile 1** | | | |
| Treatment (in standard deviations) | 0.22*** | 0.20*** | 0.22*** |
| | (0.05) | (0.04) | (0.04) |
| Treatment (coefficient in points) | 0.16*** | 0.14*** | 0.15*** |
| | (0.04) | (0.04) | (0.03) |
| **Quartile 2** | | | |
| Treatment (in standard deviations) | 0.11* | 0.12*** | 0.14*** |
| | (0.06) | (0.05) | (0.04) |
| Treatment (coefficient in points) | 0.07* | 0.07** | 0.08*** |
| | (0.04) | (0.03) | (0.03) |
| **Quartile 3** | | | |
| Treatment (in standard deviations) | 0.16** | 0.18*** | 0.19*** |
| | (0.06) | (0.05) | (0.04) |
| Treatment (coefficient in points) | 0.06* | 0.07** | 0.08*** |
| | (0.04) | (0.03) | (0.03) |
| **Quartile 4** | | | |
| Treatment (in standard deviations) | 0.09 | 0.08 | 0.09 |
| | (0.08) | (0.06) | (0.05) |
| Treatment (coefficient in points) | 0.06 | 0.05 | 0.05 |
| | (0.04) | (0.03) | (0.03) |
| Equation y intercept | 0.53 | 0.11 | 0.41 |
| **Goodness of Fit** | | | |
| F(1, 264) | **83.17** | **130.50** | **34.18** |

| | | | |
|---|---|---|---|
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.23** | **0.40** | **0.47** |
| Root MSE | 0.58 | 0.52 | 0.49 |
| Observations | 2800 | 2800 | 2800 |
| **Panel D: Effects by Gender** | | | |
| **Boys** | | | |
| Treatment (in standard deviations) | 0.22*** | 0.22*** | 0.21*** |
| | (0.06) | (0.04) | (0.04) |
| Treatment (coefficient in points) | 0.13*** | 0.12*** | 0.13*** |
| | (0.03) | (0.03) | (0.02) |
| **Girls** | | | |
| Treatment (in standard deviations) | 0.08* | 0.10*** | 0.13*** |
| | (0.05) | (0.04) | (0.03) |
| Treatment (coefficient in points) | 0.05 | 0.05 | 0.06*** |
| | (0.03) | (0.02) | (0.02) |
| Equation y intercept | 0.01 | 0.02 | 0.26 |
| **Goodness of Fit** | | | |
| F(1, 264) | **6.38** | **235.59** | **33.08** |
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.02** | **0.39** | **0.46** |
| Root MSE | 0.66 | 0.52 | 0.49 |
| Observations | 2800 | 2800 | 2800 |
| Baseline controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. All scores are expressed in standard deviations with respect to the control group.

**Table 4: Results by Geographic Location**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Urban areas** | | | |
| Treatment (in standard deviations) | 0.22** | 0.16** | 0.12** |
| | (0.09) | (0.06) | (0.05) |
| Treatment (coefficient in points) | 0.13*** | 0.10*** | 0.16** |
| | (0.05) | (0.04) | (0.07) |
| **Rural areas** | | | |
| Treatment (in standard deviations) | 0.12** | 0.14*** | 0.18*** |
| | (0.05) | (0.04) | (0.03) |
| Treatment (coefficient in points) | 0.07** | 0.08 | 0.07** |
| | (0.03) | (0.03) | (0.03) |
| Equation y intercept | -0.02 | 0.01 | 0.04 |
| **Goodness of Fit** | | | |
| F(1, 264) | **8.12** | **230.94** | **33.84** |

| | | | |
|---|---|---|---|
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.02** | **0.39** | **0.46** |
| Root MSE | 0.66 | 0.52 | 0.49 |
| Observations | 2800 | 2800 | 2800 |
| Baseline controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

**Table 5: Results by School Type**

| | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Central school** | | | |
| Treatment (in standard deviations) | 0.1 | 0.07 | 0.05 |
| | (0.09) | (0.06) | (0.05) |
| Treatment (coefficient in points) | 0.06 | 0.04 | 0.02 |
| | (0.05) | (0.04) | (0.03) |
| **Peripheral school (treatment in s.d.)** | | | |
| Treatment (in standard deviations) | 0.17*** | 0.18*** | 0.21*** |
| | (0.05) | (0.04) | (0.03) |
| Treatment (coefficient in points) | 0.10*** | 0.11*** | 0.13*** |
| | (0.04) | (0.03) | (0.02) |
| Equation y intercept | -0.03 | -0.003 | 0.25 |
| **Goodness of Fit** | | | |
| F(1, 264) | **5.08** | **227.93** | **34.03** |
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.02** | **0.39** | **0.46** |
| Root MSE | 0.66 | 0.52 | 0.49 |
| Observations | 2800 | 2800 | 2800 |
| Baseline controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

**Table 9: Effects of the Educational Level and Language of the Student's Household Head**

| Dependent Variables: | Mathematics | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel B: Effects—Language as Reported by Teacher** | | | |
| **Speaks more Guaraní** | | | |

| | | | |
|---|---|---|---|
| Treatment (in standard deviations) | 0.18*** | 0.15*** | 0.16*** |
| | (0.05) | (0.03) | (0.03) |
| Treatment (coefficient in points) | 0.12*** | 0.11*** | 0.12*** |
| | (0.03) | (0.03) | (0.02) |
| **Speaks more Spanish** | | | |
| Treatment (in standard deviations) | 0.19*** | 0.16*** | 0.17*** |
| | (0.07) | (0.04) | (0.04) |
| Treatment (coefficient in points) | 0.11*** | 0.10*** | 0.09*** |
| | (0.04) | (0.03) | (0.03) |
| **Speaks both equally** | | | |
| Treatment (in standard deviations) | 0.07 | 0.14** | 0.18*** |
| | (0.09) | (0.06) | (0.04) |
| Treatment (coefficient in points) | 0.03 | 0.02 | 0.05 |
| | (0.05) | (0.05) | (0.04) |
| Equation y intercept | 0.13 | 0.08 | 0.33 |
| **Goodness of Fit** | | | |
| F(1, 264) | **20.55** | **171.47** | **32.81** |
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.05** | **0.39** | **0.46** |
| Root MSE | 0.65 | 0.52 | 0.49 |
| Observations | 2799 | 2799 | 2799 |

**Panel C: Effects—Language as Reported by Head of Household**

| | | | |
|---|---|---|---|
| **Speaks more Guaraní** | | | |
| Treatment (in standard deviations) | 0.12** | 0.14*** | 0.17*** |
| | (0.06) | (0.05) | (0.04) |
| Treatment (coefficient in points) | 0.07* | 0.07** | 0.08*** |
| | (0.04) | (0.03) | (0.02) |
| **Speaks more Spanish** | | | |
| Treatment (in standard deviations) | 0.19* | 0.15** | 0.14** |
| | (0.10) | (0.07) | (0.06) |
| Treatment (coefficient in points) | 0.11** | 0.09** | 0.08** |
| | (0.05) | (0.04) | (0.04) |
| **Speaks both equally** | | | |
| Treatment (in standard deviations) | 0.17*** | 0.17*** | 0.19*** |
| | (0.06) | (0.05) | (0.04) |
| Treatment (coefficient in points) | 0.11*** | 0.10*** | 0.12*** |
| | (0.03) | (0.03) | (0.02) |
| Equation y intercept | 0.00 | -0.01 | 0.25 |

| | | | |
|---|---|---|---|
| **Goodness of Fit** | | | |
| F(1, 264) | **13.14** | **163.69** | **29.82** |
| Prob > F | 0.00 | 0.00 | 0.00 |
| r2 | **0.04** | **0.39** | **0.46** |
| Root MSE | 0.65 | 0.52 | 0.49 |
| Observations | 2451 | 2451 | 2451 |
| Baseline controls | No | Baseline | Yes |

*Note:* Estimated standard errors clustered at the school level are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. Test scores expressed in standard deviations with respect to the control group.