IDB WORKING PAPER SERIES Nº IDB-WP-905

Tailoring Instruction to Improve Mathematics Skills in Preschools

A Randomized Evaluation

Francisco Gallego
Emma Näslund-Hadley
Mariana Alfonso

Inter-American Development Bank
Social Sector

April 2018

Tailoring Instruction to Improve Mathematics Skills in Preschools

A Randomized Evaluation

Francisco Gallego
Emma Näslund-Hadley
Mariana Alfonso

April 2018

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Tailoring instruction to improve mathematics skills in preschools: a randomized evaluation / Francisco Gallego, Emma Näslund-Hadley, Mariana Alfonso.
p. cm. — (IDB Working Paper Series ; 905)
Includes bibliographic references.
1. Mathematics-Study and teaching (Preschool)-Peru. 2. Education, Preschool-Peru. 3. Early childhood education-Peru. 4. Child development-Peru. I. Näslund-Hadley, Emma. II. Alfonso, Mariana. III. Inter-American Development Bank. Education Division. IV. Title. V. Series.
IDB-WP-905

http://www.iadb.org

TAILORING INSTRUCTION TO IMPROVE MATHEMATICS SKILLS IN PRESCHOOLS:

A RANDOMIZED EVALUATION

Francisco A. Gallego[*] Emma Näslund-Hadley[†] Mariana Alfonso[‡§]

Abstract

Tailoring instruction to each student's needs can produce significant learning gains. However, few programs have successfully implemented this approach. In this paper, we present the results of a randomized evaluation of a program that uses an inquiry with an individualized scaffolding approach to teach Mathematics to preschoolers in Peru. Our results suggest that the program improves overall Mathematics outcomes, and that it has stronger impacts on students in the lower quintiles of the Mathematics outcomes distribution and on students whose teachers have university degrees. The effect on the content areas where the program was implemented more intensively persists even one year after the program ended. We find no evidence of differential effects by gender, language-spoken at home, and proxies for SES, in contrast with results from previous research that suggest Mathematics programs are biased along gender and socioeconomic lines.

Keywords: Education, Mathematics, Early Childhood Development, RCT, Scaffolding.
JEL Classification: I21, I28, O15.

---

[*] Francisco A. Gallego, Pontificia Universidad Católica de Chile and J-PAL. email: fgallego@uc.cl
[†] Emma Näslund-Hadley, Inter-American Development Bank. e-mail: emman@iadb.org
[‡] Mariana Alfonso, Inter-American Development Bank. e-mail: marianaa@iadb.org
[§] Mimate is a program designed and implemented by Instituto Apoyo in coordination with the Ministry of Education of Peru (MINEDU) and the Inter-American Development Bank (IDB). We thank Inés Levy, Sergio de Marco, and Alejandro Saenz for excellent research assistance. We benefited from comments from Samuel Berlinski, Jessica Goldberg, Elizabeth Spelke, and audiences in seminars at Harvard, PUC-Chile, and the Workshop on "Equity: Contributions from research in Education, Economy, Psychology and Neurosciences" held at PUC-Chile in March 2016. Innovations for Poverty Action (IPA) assisted with the sampling, design field work, and statistical analysis of the data. We are also grateful to the Inter-American Development Bank, J-PAL, and the Government of Japan through the IDB JPO Trust Fund for financial support. AEA RCT Registration Number: AEARCTR-0000365. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.

1. **Introduction**

The relevance of early childhood development and pre-school education for subsequent development of cognitive and non-cognitive abilities and economic and social performance is well known (Cunha and Heckman, 2007). A key mechanism to explain this effect is related to dynamic complementarities, in which skills developed at early ages complement the future production of new skills. This is also true for the development of mathematics skills, as mathematical thinking is cognitive foundational (Clements and Sarama, 2009, 2011). Poor mathematics performance throughout primary school can be traced to weaknesses in pre-primary number competencies (Gersten and Flojo, 2005; Malofeeva et al. 2004; National Mathematics Advisory Panel, 2008). Moreover, several studies document that mathematics skills acquired in kindergarten predict academic performance in primary education better than early reading skills, verbal and spatial cognition, or even measures of attention span, memory, and social skills (Duncan et al., 2007, Romano et al., 2010, Jordan and Locuniak, 2009, Pianta et al., 2008a, Aunola et al., 2006). Still other studies find impacts of early mathematics learning on adult life outcomes (Geary et al., 2013).

However, it is less clear which specific pedagogical approaches can improve learning for preschoolers and affect subsequent learning (Duncan and Magnuson, 2013; Yoshikawa et al., 2013; Berlinski and Schady, 2015; Clements and Sarama, 2011). Recent research suggests that changes in pedagogy—with the same stock of teachers—in primary schools can lead to significant improvements in learning outcomes, especially in the case of programs focused on teaching to each student's needs—the "teaching at the right level" approach (e.g., Banerjee et al., 2016). One way to implement this idea is the *scaffolding* approach. Scaffolding is an adaptive interactive strategy that recognizes the current capacities of children (trainees) and guides them to further learning without creating frustration. Several papers document that some of the most promising recent models of parent-child, teacher-child, and parent-teacher-child relationships involve attachment and scaffolding as major determinants of child learning (Heckman and Mosso, 2014). Activities are tailored to the individual child's ability level, so they are neither too hard nor too easy. This approach keeps the child in the "zone of proximal development", which is the level of difficulty at which the child can learn the most (Heckman and Mosso, 2014). In the same way that a scaffold is used as a temporary structural support during building construction, scaffolded instruction provides targeted and temporary support to help students develop new skills and abilities (Englert et al., 1991).

In this paper we study the Mimate program, which uses an inquiry and individualized scaffolding approach to teach the basic elements of numbers and shapes to preschoolers in Peru. The program seeks to change the traditional vertical teaching model of memorization and repetition into a horizontal, student-centered model in which students progress at their own rate.[1] The program includes sessions, implemented during the regular schedule of Peruvian preschools, in which the teacher splits up the children into small groups or pairs for activities. The lesson plan proceeds with mathematics challenges, gradually progressing from very basic to advanced, in which each task prepares the student to tackle the next one. Mimate emphasizes the development of early mathematics competencies in two main areas: numerical literacy and understanding shapes. Each student also receives a personal box with hands-on teaching materials to encourage exploration of mathematical ideas, shapes, measurements, numbers, and patterns. All items are suggested to be kept in the "Mimate corner" of the classroom so that children can play with these toys and tools at free hours during the day. The program also includes twice-a-month formative assessments with a simple 5-minute round of flash cards between the teacher and individual students. Based on the students' answers, the teacher then knows which skills the students need to practice and can direct them to an appropriate activity. In addition, there is an initial training period for the teachers before the program starts, and in-class visits from teacher assistants during the school year to ensure the quality of teacher-student interactions. The cost of the program is about $37 per student (in 2013 US dollars) in addition to the regular educational expenses.[2]

The randomized controlled trial (RCT) was implemented in schools located in and around three cities of Peru: Huancavelica, Angares, and Ayacucho. We implemented a stratified randomization with 54 treatment schools and 53 controls schools, stratifying by cities and urban/rural area. The baseline information we collected suggest that the randomization was successful, as most variables were balanced across treatment and control groups.

We find that, at the end of the program, the intervention increased mathematics outcomes for children that attended the treated schools by 0.10 standard deviations (σ hereafter) for items related to numeracy and by 0.12σ for items related to understanding shapes. In addition, when one looks at specific items, we find that the impacts are bigger for geometric shapes (impact of 0.20σ), number selection (impact of

[1] Algan et al. (2013) document the presence and consequences for social capital of teaching practices that emphasize vertical teaching methods in contrast with group work and a more horizontal teaching practice in several countries of the world.
[2] This is the cost per student to scale-up Mimate. The pilot cost was $150 per student.

0.18σ), additive composition (impact of 0.15σ), and oral counting (impact of 0.11σ). The effects in these dimensions are statistically significant even when correcting the p-values for multiple hypotheses testing.[3]

We also study the potential existence of heterogeneous effects along several dimensions related to characteristics of the students and the schools. First, using quantile regressions, we find that the impact of the program is stronger for children in the lower quantiles of the learning distribution. This contrasts with previous evidence that finds stronger effects for students with good mathematics performance (Clements and Sarama, 2011; Duncan and Magnuson, 2013). This result suggests that the inquiry-based scaffolding approach may work better for students with lower abilities, who receive more attention than when taught using more traditional pedagogical approaches. Second, we study interaction effects considering several observable characteristics of the students and the schools, such as: the student's gender, language and socioeconomic status, location of the school, class size, and teacher education. We do not find that the treatment has statistically heterogeneous effects on these dimensions except for teacher education: our results indicate that when the teacher who implemented the program has a university degree, the average effect of the program increases to about 0.15σ (and the effect of the program for teachers without a university degree is not different from 0). This suggests that there is a complementarity between the teacher's human capital and the program, which probably indicates that a program like this demands a certain level of teacher human capital to successfully work with inquiry and scaffolding.

Next, we study the medium-term treatment effects, using information on the same outcomes for the same students one year *after the program ended*. Our results imply that the effects decrease in magnitude after one year. The treatment effects for the overall mathematics test and for the numeracy section are not different from 0, and treatment effects for the understanding shapes section are equal to 0.06σ. These results resemble most of the findings from the literature on preschool interventions: significant, short-term effects on learning skills that decrease after the programs end (Duncan and Magnuson, 2013). The only item for which we find a large and statistically significant medium-term effect is geometric shapes, with an impact of 0.16σ.[4] As we discuss below, the larger impacts we find in this curricular area may be a consequence of the intensity of implementation.

[3] To correct p-values we use a family-wise error correction approach, as suggested by Anderson (2008). We use this approach in all the results for which we test effects on different dimensions.
[4] Correcting for multiple hypothesis testing.

In terms of medium-term heterogeneous effects, and mirroring our short-term results, we find that the only statistically significant and positive interaction effect is the one related to having a Mimate teacher with a university degree. Our quantile regressions suggest that, in the medium-term, the positive effects we estimate for understanding shapes seem to be concentrated in the students in the upper part of the learning outcomes distribution. This result may suggest that, once the program ends and students continue in first grade exposed to regular pedagogical approaches, those with higher learning outcomes are able to continue learning at a faster pace.

Finally, we study whether the program has an impact on other dimensions of the learning process. On one hand, there may be potential positive externalities of a program like Mimate if the new abilities that the students acquire also affect child development and learning in other areas. On the other hand, there may be negative effects if the effort to improve mathematics outcomes displaces resources (such as time, teacher effort or even student interest) from other activities and learning dimensions. We do not find evidence of indirect effects using results from the Raven test of general cognitive ability and from an early literacy test.

To facilitate the interpretation of the results, we also collected information on the program's actual implementation using administrative data, class observations, and surveys applied to teachers and parents. The process evaluation suggests that, on average, 66% of the 86 planned 45-minute lessons were implemented. This partial implementation was mainly due to a national teacher strike that suspended up to three months of class time in some schools. This may explain our findings, given that sessions on number sequence, quantities, and patterns came at the end of the program and were covered less intensively, in contrast to the sessions on shapes and figures that were covered at the beginning of the school year. In fact, our process evaluation finds that while 82% of the sessions planned to cover figures were implemented (equivalent to about 30 sessions), just 57% of the sessions related to numbers were implemented (equivalent to about 25 sessions).[5] This may explain both our short- and medium-term results. In addition, results from the class observations and the teacher questionnaires suggest that the program affected students' and teachers' classroom behavior, the teachers' beliefs about their students, and the teachers' ability to meet the objectives and to teach mathematics (all of this with respect to the

[5] This is a very coarse classification between numbers-related and shapes-related sessions, as the program seems to influence both dimensions in an interlinked way. However, we want to illustrate the differences between dimensions in sessions covered by the program.

teachers in the control group). In contrast, teachers in the treated group do not report significant differences in terms of the availability of resources and mathematics topics covered in comparison to the control group. This suggests that the intervention operated mainly through changing teaching practices more than through increases in resource availability.

Although there is a considerable literature exploring the causal impacts of preschool programs (see review in Almond and Currie, 2010) and exploring the effect of pedagogical innovations (see review in Kremer et al., 2013), only few studies have conducted randomized controlled trials to study the effects of implementing pedagogical innovations among preschoolers in the developing world (He et al., 2009; Naslund-Hadley et al., 2014).[6] Thus, this paper makes several contributions to the existing literature. First, this study represents one of the few randomized evaluations of mathematics programs and the only randomized evaluation that combines a relatively large sample of schools and students and follows up the students after the program ends in order to identify medium-term effects. Second, this paper contributes to the research on inquiry-based and individualized instruction that allows students to solve problems, develop explanations, and communicate ideas, while adapting the teaching process to the learning pace of each student. This is related to both the "teaching at the right level" (TaRL, hereafter) (Banerjee and Duflo, 2010; Banerjee et al., 2016) and the scaffolding (Heckman and Mosso, 2014) empirical literatures.[7] Regarding the TaRL literature, the Mimate program is closer to the studies that analyze the impact of classroom-based interventions that try to change the pedagogical practices of the current stock of teachers (He et al., 2008; Banerjee et al., 2010, Barrow et al., 2009, Muralidharan et al., 2017). Regarding scaffolding, while several papers have studied the impact of these types of programs in different contexts (Lynch and Kim, 2017; Kim et al., 2017), to our knowledge this study represents the first randomized evaluation of a scaffolding approach in mathematics instruction in developing countries. Third, we use evidence from both a process and an impact evaluation to suggest that the part of the program that was implemented as planned—that related to instruction about geometric shapes—tends to have larger effects in the short-run and some persistent effects in the medium-run. Moreover, the information collected from

[6] We do not discuss in detail the literature on the effects of early childhood interventions in developed and developing countries that track individuals from early childhood into adulthood and show that children brought up in a more favorable early environment are healthier and taller, have higher cognitive ability and educational attainment, and earn significantly higher wages (see Paxson and Schady 2010; Stith, Gorman and Choudhury 2003; Liddell and Rae 2001; Walker, et al. 2005; Gertler, et al. 2014; Havnes and Mogstad, 2011).

[7] Both approaches are related in the sense that they try to implement specific ways of teaching that adapt to each student level. However, a relevant difference among them is that in the scaffolding approach the specific way of adapting to each student is more focused in the relationship and interaction between the teacher and the student.

teacher surveys suggest that the program effects are derived from changes in pedagogy rather than from increases in resources or the mathematics topics covered in class. This is an important contribution to the literature, which typically does not present detailed information about actual program implementation.

The remainder of the paper is structured as follows. Section 2 briefly presents Peru's educational context, and a description of the Mimate program and its relationship with the literature on mathematics preschool programs. Section 3 describes the research design and the data used in the paper. Section 4 describes the implementation of the program using administrative data, class observations, and answers to a teacher survey applied after the program was implemented. Section 5 discusses the main results, including both short- and medium-term effects. Finally, Section 6 provides a summary of our findings and draws some policy implications.

2. **Context and the Mimate Program**

a. Education in Peru

Access to education in Peru is high, with net school enrollment rates in primary and secondary education above 95% and 75%, respectively.[8] Although learning outcomes have also improved considerably in the last decade, they are still low compared to countries with similar income levels and present steep socioeconomic gradients. At 15 years of age, according to the 2015 PISA report, Peruvian students ranked 62 in mathematics, 63 in reading, and 64 in science out of the 70 countries that participated. Moreover, Peru is a country that displays one of the largest performance gaps in PISA between low-income and high-income students (Bos et al., 2016). Similarly, results from Peru's 2nd grade national census evaluation show a difference of about one standard deviation between children in the richest and poorest quintiles (Berlinski and Schady, 2015). Schady et al. (2015) document similar differences for six-year olds in urban and rural contexts of Peru. Even at five years of age, Peruvian children score on average 15 percentage points lower on the Raven test of cognitive ability than five-year olds in the average of countries for which data is available (See Table 1). This difference is sizeable, as it is equivalent to about 60% of a standard deviation observed for five-year olds in the United Kingdom.

[8] Gross enrollment rates are for 2013. See IADB-CIMA (2018), accessible on https://www.iadb.org/es/cima

The Peruvian Ministry of Education (MINEDU) has recognized pre-primary school as a priority to improve educational outcomes. As a result, from 2007 to 2011, MINEDU increased pre-primary education spending per student by 70%, compared to increases of 60% and 46% in primary and secondary education respectively (ESCALE, MINEDU). In parallel, the rate of attendance to preschool for children ages 3 to 5 years increased from 53% to 75% in the 2001-2012 period, with increases for children living in both rural and urban areas (attendance increased from 53% to 75% in urban areas and from 44% to 66% in rural areas during the same period). Moreover, Berlinski and Schady (2015) report that the gap in enrollment between students from the richest and poorest quintiles dropped from 36% to 12% between 2000 and 2013.

This effort to expand access to pre-primary education, however, was not initially accompanied by programs aimed at changing the pedagogical practices as a vehicle to improve learning outcomes. Thus, the status-quo in terms of pedagogical practices in pre-primary education was one of memorization and repetition using a model of vertical pedagogy, in which all the students in the classroom received the same materials and tended to repeat, at the same time, what the instructor did—e.g., all students would sing at the same time or repeat the number the instructor mentioned in front of the class (APOYO, 2012).

b. The Mimate Program

Mimate was designed by pedagogical experts with the relevant literature on early education kept closely in mind. The program is delivered in 45-minute sessions to fit smoothly into the daily schedule of Peruvian preschools, and includes a workbook, formative assessments, and visits from teacher assistants to ensure the quality of teacher-student interactions.

The salient characteristic of Mimate is the use of inquiry with scaffolding techniques. Inquiry-based learning, or co-operative learning as Vygotsky (1978) called it, is based on the idea that social interaction is essential for learning. Inquiry-based learning tends to rely heavily on scaffolding to guide learners through complex tasks and keep the student engaged in what Vygotsky (1978) described as the "zone of proximal development," the range of concepts between what learners can do on their own and what can be achieved in collaboration with instructors or peers (Ellis and Worthington, 1994). Scaffolding can be achieved using soft or hard approaches (Saye and Brush, 2002). Soft scaffolding is human support provided by a teachers or peers that helps students meaningfully participate in the performance of actions. Teachers who provide soft scaffolding must remain cognizant of students' stages of learning in order to provide just the right amount of support at the right time to each student (de Grave et al. 1999; Hogan and Pressley 1997; Lepper et al. 1997). Hard scaffolds are computer or paper-based materials that anticipate the student's needs during the unit by following a pedagogical flow that builds upon concepts incrementally (Saye and Brush, 2002). Ideally, both soft and hard scaffolding measures can occur simultaneously to allow teachers to keep an entire classroom engaged with the material. Mimate uses both elements.

Individualized instruction requires teachers to know exactly where each child stands in terms of her understanding of the material.[9] This is no simple task in classrooms of up to 30 students per teacher and without access to technology. Mimate developed a solution that makes accurate, formative assessments possible with a simple 5-minutes round of flash cards between the teacher and each individual student.[10] Based on the student's answers, the teacher then knows which skills the student needs to practice and can direct her to an appropriate activity.

The nature of the program also needed teacher training, which included two different dimensions. Before the program started, teachers participating in the program received three training sessions in which they received information on the objectives of the program and training on the program's sessions and materials. 83% of the teachers attended these sessions. The second component of training support was in-class training. For this, teacher coaches observed the sessions taught by the teachers and gave advise on how to improve the application of the program. These sessions took place during the school year (from March to November). On average, teachers received six visits.[11]

In terms of mathematics contents, Mimate is motivated by research that suggests preschoolers can learn mathematics and, moreover, most children at the preschool age actually enjoy the basic mathematics challenges (Seo and Ginsberg, 2004; Clements and Sarama, 2008). Similarly, research also shows that

---

[9] This is related to the literature that emphasizes the importance of the quality of the interactions between students and teachers in order to attain good learning outcomes (see Berlinski and Schady, 2015 and the references therein).

[10] Recently, sophisticated computer programs have become popular tools to test students and recommend further exercises based on their specific capabilities (Barrow et al., 2009; Muralidharan et al., 2017). However, these types of programs were not feasible to implement in the Peruvian context as only 7% of schools had internet access and computer skills among teachers were sparse.

[11] In addition, during the final part of the implementation of the program (between October and December), participating teachers were invited to attend lesson study sessions in which they could share their experiences with other teachers, model classes. There were four of these sessions and, on average, 63% of the teachers attended these sessions.

preschool mathematics programs aimed at children from low socioeconomic status can have significant impacts on their mathematics outcomes (Ramani and Siegler, 2011).[12]

The Mimate program emphasizes the following four areas of mathematical development:

a) *Numerical literacy: number sequence and number aspects.* Children learn the variety of purposes for numbers: to count one-by-one, to order objects, to measure size or time, etc. The two basic types of "number words" (i.e., one, two, three, etc. and first, second, third, etc.) are also practiced.
b) *Numerical literacy: capturing structure.* Children practice the rule of cardinality, that the last number spoken when counting signifies the total number. When counting is understood, children are stretched to consider numbers in small groups and learn how to combine and separate groups mentally without relying on counting. This catalyzes quick and efficient mental calculus later on.
c) *Understanding shape: variety of geometric shapes.* Children play with various shapes (circles, squares, rectangles, triangles and rhombuses), different types of lines (straight or curved, closed, or open). They are challenged to visualize points, planes, and three-dimensional spaces.
d) *Understanding shape: training fine motor skills.* Children construct shapes with clay, draw shapes with pencils, and learn symmetry by folding or cutting paper. Neuroscience has demonstrated that knowledge of form comes through experiential learning with the hands in one channel and language in the other. The activities hone the coordination of fingers while reinforcing the lessons of earlier modules.

3. **Research Design and Data**

A randomized controlled trial employing randomization at the school level forms the basis of the study design. We randomly selected 53 treatment schools and 54 control schools from Huancavelica, Angares, and Ayacucho. The treatment group would adopt the Mimate program in their preschool classrooms, while the control group would not receive the intervention and continue using traditional pedagogical practices (as we discussed above, generally this means teaching practices based on frontal instruction, memorization

[12] It is well known that low-income children enter preschool with fewer number-related experiences—at home, in their communities, and in their early care centers—than their middle-income counterparts and that these early learning opportunities set middle-income children at an advantage (Clements and Sarama, 2008).

and repetition). The randomization considers the existence of six strata that come from the combination of the school location (urban/rural) and the city in which the school is located.

The primary data used in this paper were collected through the baseline and two follow-ups. Baseline data were collected in March 2012. A first follow-up was collected at the end of the school year in December 2012 (referred to as “short-term follow-up” in the paper). We collected a second follow-up in December 2013, one year after the treatment was completed and when the students were at the end of their first grade of primary education (referred to as “medium-term follow-up” in the paper). In total, 2,400 children participated in the baseline and short-term data collections, while 2,416 participated in the baseline and medium-term data collections.

The general format of these data collection processes was very similar, consisting of the application of student tests in the baseline and two follow-ups and the application of parent and teacher surveys in the baseline and first follow-up. To measure the development of mathematics skills, we applied preschool-adapted versions of the “Early Grade Mathematical Assessment” (EGMA) originally developed by the Research Triangle Institute International. The tests measure various abilities related to mathematics with the following exercises: Comparing quantity, basic shape recognition, advanced shape recognition, basic object counting, advanced object counting, number selection, advanced number selection, fine motor skills, symmetry, shape sequence, number sequence (clock), number sequence (calendar), additive composition, geometric shapes, and addition and subtraction word problems. Appendix 1 provides a detailed description of these items. In the paper we present results of the items included in EGMA at three different levels of aggregation: (i) an overall index that includes all the items in EGMA, (ii) two indices by curricular area, one for items related to numerical abilities and the other for items related to shapes, and (iii) measures of the development of each ability included in EGMA. The preschool version of EGMA was originally developed for Paraguay and was validated to align the test with the Peru preschool curriculum and pretested for use in Ayacucho and Huancavelica (CPAL, 2012; IPA, 2012). The most important adjustment was the inclusion of additional items to assess fine motor skills based on Hammil and colleagues (1995). In addition to mathematics tests, we also applied instruments to assess the development of non-mathematics outcomes: the Raven test of general cognitive ability and a test of early literacy skills test.[13]

[13] Early literacy skills were measured through a preschool version of the Early Grade Reading Assessment (EGRA).

The questionnaires applied to teachers and parents collected information about the child's classroom and home experience. We use the information from these surveys for two purposes. First, to study treatment heterogeneous effects (such as classroom size, access to educational materials, teacher's education, mother's education, dominant language at home, etc.). Second, the information from these surveys allow us to understand the mechanisms through which the program may have affected learning outcomes.

In addition, to help with the interpretation of the results, we collected process information about actual program implementation in all treatment schools. Lastly, between the months of October and December of 2012 we also visited 44 randomly selected schools (37 treatment and 7 control schools) to verify the intensity of treatment (teacher and student attendance records and level of implementation of different Mimate lessons) and to explore any indication of contamination. In addition, to observe teacher-student interactions and student-student interactions, 10 schools were randomly selected from each group. In these 20 schools, mathematics classes were videotaped and analyzed using the CLASS (Class Assessment Scoring System) rubric, which codes teaching practices in three areas: emotional support (i.e., generating a respectful environment and listening to children), classroom organization (i.e., managing time and keeping control of students), and teaching support (i.e., developing concepts thoughtfully and reinforcing student learning). With respect to this third category, the instrument provides a general indication of the teaching approach used but does not detail the precise pedagogical practices used.[14]

4. **Program Implementation**

This section presents information to characterize several dimensions of Mimate's implementation. This is important for three reasons. First, it allows us to know fidelity of treatment implementation. Second, a comparison with the pedagogical practices in control schools allows us to understand the intensity of the different components that may be affected by the program. Third, it also allows to understand the status-quo's pedagogical practices (those of the control group). This information can provide us with useful insights to interpret the program's quantitative results. To analyze program implementation, we use information from three different sources that we discuss below.

[14] See Pianta et al. (2008b) for a description and Araujo et al. (2016) and Cruz-Aguayo (2015) for applications in Latin American countries. For cost reasons, we were not able to other instruments (for example, the TIMSS video study) that would have allowed us to measure the precise pedagogical practices used by teachers.

a. *Process Information*

Concurrently to the experimental evaluation, we collected data to produce rough measures of the program's actual implementation both in terms of quantity and quality.

First, we collected administrative data on all the sessions implemented by each school. Table 2 presents a general description of how the program was planned and implemented from March to December. The program planned 86 sessions in total. The administrative data implies that, on average, 66% of the sessions were implemented. It is important to highlight that during the school year 2012, a teacher strike suspended up to three months of classes in some schools. In total, 87% of teachers in the sample participated in the strike at some point in time. This rate was higher in rural areas (95%) than urban areas (83%).

The partial implementation of the program implied that the actual coverage of the topics planned to be taught in the last three months of the school year was smaller than that for topics planned for the beginning of the year. For instance, while all the topics planned for March to July were implemented by all schools, none of the topics planned for November and December were covered. Given the intended order of the sessions, this partial implementation means that actual coverage of the numerical literacy and the understanding shapes parts of the program was significantly different. In fact, while 82% of the sessions directly related to shapes were implemented, just 57% of the sessions directly related to numeracy were covered. This is an important finding to consider in our empirical analyses.

To analyze whether there are some systematic implementation patterns related to the schools' observable characteristics, in Table 3 we present a regression in which we study the correlation between the percentage of sessions covered in each treatment school and students' (baseline mathematics scores, gender, language, and socioeconomic status) and schools' (location, average class size, and teacher education) observable characteristics. Columns (1) to (7) show the relation between the percentage of completed sessions and each variable individually, and column (8) includes all the variables. Our models do not successfully predict the number of implemented sessions, as all the variables combined can just explain about 22.5% of the variance in the number of sessions (column 8). In terms of individual significance, we find that schools with higher test scores at the baseline, with more Spanish speaking students, with more students coming from high SES households, and located in urban areas implemented more sessions. However, in all the cases, the size of the impact, while statistically significant, is not economically relevant. For instance, the estimated coefficients imply that a school with students with mathematics scores one standard deviation above the average school implemented just 5 percentage points, or close to 3, more sessions than the number of sessions completed on average (57). Similarly, rural schools decrease the percentage of implemented sessions by just 5.7 percentage points with respect to urban schools and schools with only bilingual students reduced the share of implemented sessions by just 8.1 percentage points (close to 5 sessions) with respect to schools with just Spanish speaking students. Moreover, when including all the variables together (column 8), none of them has a correlation that is statistically different from 0—which could probably be a consequence of the high collinearity among the right-hand side variables. In all, these results suggest that while some variables seem to be correlated with the percentage of completed sessions, none of these variables seem to have a first order effect on the program's actual implementation.

b. Class observations

Next, to study the quality of the program's implementation, we use class observations from a sample of 44 schools (37 from the treatment group and 7 from the control group). The idea was to identify a number of relevant pedagogical dimensions and see whether actual implementation was consistent with the program's objectives.

We now report findings from the monitoring of the intensity of the treatment. All the treatment schools have the "Mimate corner" in the classroom and in 78% of the cases the actual set-up of the corner was consistent with the design suggested by the program. The board games provided by Mimate were available to students in 92% of the treatment schools. In 100% of the treatment schools, teachers were using the Mimate materials as suggested by the program. These results suggest that implementation fidelity was very high and that teachers were using the Mimate program as designed.

We identify several interesting results when comparing treatment and control schools using data from class observations. First, in the area of classroom organization, the Mimate classes were 38% more likely to be "prepared and structured with a clear objective" than control classes. Teachers in the control group, on the other hand, were many times observed improvising with activities during the lesson.

Second, in the area of emotional support, teachers in the treatment group explained mathematics patiently more often than their colleagues in the control group (95 vs. 71%). Moreover, when students committed mistakes, Mimate teachers more consistently encouraged, in a friendly manner, their students to try activities multiple times (in 95% of the cases), compared to the control group (with 71% of cases).

Third, in the area of teaching support, 94% of the teachers in the treatment schools allowed their students to explore and manipulate the Mimate materials but there were some differences between urban (100% of the teachers allowed their students to manipulate the materials) and rural schools (78% of the teachers allowed their students to manipulate the materials). Consistent with the program's inquiry-based approach, in most cases teachers in treatment schools were always or often allowing their students to discover by themselves the objectives of the activities (an average of 3.5 in a metric that goes from 0 to 4, with 1=never, 4=always). In 100% of the treated schools, boys and girls were equally included in classroom activities. Furthermore, Mimate teachers were observed to be "paying attention to those students who do not understand well and explaining with patience" at a higher rate (95%) than the control group (63%). Thus, consistent with the scaffolding approach, the Mimate program seems to have helped teachers use class time more efficiently to reach to students that lag behind and adopt a more patient attitude.

The CLASS rubric allows us to quantify the above findings along the three areas described above[15]: (i): the average rate of emotional support was higher in treatment schools (5.32 compared with 4.78 in control schools, and the difference is equivalent to about 0.72σ), (ii) the level of classroom organization was higher in the treatment schools (4.47 in versus 4.30 in control schools, difference equivalent to 0.16σ), and (iii) the teaching support was also higher in the treatment schools (3.93 versus 3.79 in control schools, difference equivalent to about 0.15σ).[16] In sum, we find some differences in favor of treatment schools—with the caveat that it is difficult to make statistical inferences from these observations given the small sample size.

c. *Teachers and parents' follow-up surveys*

We use answers to surveys applied to 101 teachers in 91 schools and to 1,780 parents in 99 schools after the treatment was applied, as a source of additional information on the program's actual implementation. Given that we also include control schools in this sample, we are able not only to understand the changes created by the program but also the counterfactual situation without the treatment.[17]

---

[15] The rating scale for CLASS goes from 1 to 7, with 7 being the highest score.

[16] As a reference, Cruz-Aguayo et al. (2015) report results of the application of the CLASS instrument in classrooms in grades K-2 in Brazil, Chile, and Ecuador. Their results are similar to what we observe in our study, with relatively lower performance in the area of teaching support and similar results for emotional support and classroom organization.

[17] All the differences between the treatment and control group are statistically different from zero (using standard errors clustered at the school level).

We find several significant differences between the treatment and control groups when we analyze data from the teachers' survey. Regarding teaching mathematics, treatment teachers are more likely to agree that they have had enough time to teach all the material they were supposed to cover (71% versus 48% in the control group). They are also more likely to disagree with the statement "children get tired when learning mathematics" (32% versus 11% in the control group). This may be related to their higher likelihood of reporting to strongly agree with teaching mathematics using games that include all students in the class (55% among treated teachers versus 38% among control teachers). In terms of preparation, treated teachers are more to likely to strongly agree that they are accurately prepared to teach mathematics (23% versus 7% in the control group). They are also more likely to strongly disagree with the idea of teaching at another educational level (42% versus 17% in the control group).

Teachers also report to observe significant changes in their students. Teachers in treated schools are more like to report that: (i) most of their students behave well in class (68% versus 50% in the control group), (ii) most of their students have the capacity to learn both mathematics (46% versus 29% in the control group) and also any material they are taught (55% versus 39% in the control group), (iii) more than half of their students do not have problems at home that affect their learning (81% versus 53% in the control group), (iv) just a minority of students do not have support from their parents (15% versus 30% in the control group), (v) just a minority of students have concentration problems while in class (17% versus 32% in the control group), and (vi) most of their students have good performance in mathematics (55% versus 36% in the control group). Thus, teachers improved their opinions about the abilities, performance, behavior and the social context of their students.

A challenge that all mathematics programs face is related to the potential gender-bias that can be present when teaching the subject. Interestingly, teachers in the Mimate program have more egalitarian beliefs about what happens in the classroom and about their students. The answers to the endline survey imply that teachers in treatment schools are more likely to report that there are no differences among children of both genders in terms of: (i) discipline problems (35% versus 18% in the control group), (ii) demands for the teacher's attention (67% versus 45% in the control group), (iii) paying attention in class (49% versus 38% in the control group), (iv) following instructions in class (67% versus 39% in the control group), and (v) performance in mathematics (80% versus 54% in the control group). These results, especially the last one, suggest that the program was able to introduce a more balanced view of gender among teachers and to the teaching and to the learning of mathematics. This is important to interpret the results on learning we present below.

Finally, we also use the parent surveys after treatment to understand whether the program changed the beliefs, attitudes, homework, and involvement of parents in their children's education. We do not observe significant differences among parents in treatment and control schools in any of these dimensions. The only exception is that treatment parents are more likely to strongly agree that their children will do well on mathematics in school (41% versus 33% in the control group). This suggest that the program does not seem to have affected parenting beliefs about mathematics in a significant way.

In sum, the process information reported in this section suggests that, even when the program was only partially implemented due to teacher strikes, results from both class observations and teacher surveys suggest that the program affected several margins related to the teachers and students' behavior in the classroom in ways that are consistent with the approach of the program.

5. **Empirical Results**

We start the discussion of the empirical results by characterizing the population included in the study at baseline. The baseline tests completed by all study participants revealed a real academic disadvantage for Peruvian preschoolers compared to international standards (see Table 1). This implies that their basic mathematics competence was of concern. For instance, fewer than four in ten children could write their age and only about half of the children could count to the number ten. As for the teachers, 12% had earned a Masters or Doctorate degree, 44% possessed a university degree, and the rest had earned a tertiary-level teaching certification. On average, teachers had 14 years of experience and were responsible for a class of 24 students.

Table 4 presents an analysis of heterogeneity in mathematics outcomes across different groups using the baseline data. We present standardized outcomes (i.e., with an average of zero and a standard deviation of one) for tests that measure mathematics, cognitive ability, and early literacy skills. We identify significant gaps in learning among most dimensions: urban vs. rural students; Spanish, Quechua, and bilingual speakers; and males vs. females. This is not surprising when one considers that, when looking at the infrastructure and resources of the schools, less than half of the participating rural schools were connected to the public water system, compared to 88% of urban schools, and one in five had a computer, compared to one in two for urban schools. In this setting, rural students scored on average 0.46 standard deviations below their urban peers in the baseline mathematics assessment.

Two in three of the participating rural students spoke exclusively Quechua at home, and the rest were split evenly between bilingual and Spanish speakers. Only 2% of participating urban students spoke exclusively Quechua at home, and the remainder spoke only Spanish. The Quechua speakers performed at baseline even lower than the rural group (0.61 standard deviations below Spanish-speaking students), and the bilingual students also had lower performance compared to Spanish speaking students (0.46 standard deviations below Spanish speaking students).

In the remainder of this section, we first present the econometric models and statistical tests we use. Then, we perform an initial check of internal validity and attrition in the three data collection processes. Lastly, we present the main results for both the short-term and medium-term follow-ups and considering both average and heterogeneous treatment effects.

a. *Econometric Models and Statistical Tests*

The empirical framework for the main analyses of this paper uses observations on children who attended schools that were randomly assigned to the treatment and control groups. We estimate the following regression model:

$$Y_{ij} = \theta + \alpha T_j + X'_{ij}\beta + \varepsilon_{ij}, \tag{1}$$

where Y represents an outcome for student i who attends school j, T is a dummy variable that takes a value of 1 if the school is part of the treatment group, X is a vector of control variables, and ε is the error term that is clustered at the school level j in each regression. Then, estimates of α quantify the differences in means between students attending schools that received the Mimate program over the 2012 school year and those that did not receive it.

We include the following variables in vector X: the value of Y for student i at baseline, strata dummies, and dummies for the different enumerators. If the randomization is successful, adding these variables should not change the estimate of α and these controls should just provide more precise estimates. For instance, it is well established that tests scores at the student level present high variation levels. Therefore, controlling for baseline estimates reduces the idiosyncratic variation in the results and allows us to obtain more precise estimates (Duflo et al., 2008). We find that the differences in the estimates of α when including and excluding X in equation (1) come from differences in the precision of the estimates of α. Therefore, to save space, we only report estimates including X in the tables.

In addition, we also implement the following additional econometric exercises to analyze heterogeneous treatment effects. First, we run regressions considering interaction effects between the treatment dummy and students/schools observable characteristics of the following form:

$$Y_{ij} = \theta + \alpha T_j + \gamma T_j * W_{ij} + \eta W_{ij} + X'_{ij}\beta + \varepsilon_{ij}, \qquad (2)$$

where W corresponds to a variable that varies at the school or student level. Then, if we estimate that γ is statistically different from 0, it implies that treatment effects are different for students having different W.

Second, we use quantile regressions to estimate the effects of treatment for students in different quantiles of the outcomes distribution, following a structure similar to the one presented in equation (1). In this case, we obtain an estimator of α that is different for individuals in different quantiles of the distribution, conditional on outcomes in the baseline (as we include them in the vector of covariates).

As we include different outcomes, we may face an inference problem related to multiple hypotheses testing. That is, significant coefficients may emerge simply by chance, even if there are no treatment effects (Anderson, 2008), and we may be over rejecting the null-hypothesis of no treatment effect. Thus, in the case of tests for treatment effects on individual components of the test, we calculate adjusted p-values following the methodology suggested by Anderson (2008). This procedure corrects the p-value for the family wise error rate (the probability of making a Type I error) and uses a Westfall and Young (1993) type correction. We call these FWER p-values.

b. Balance and Attrition

Panel A of Table 5 compares outcomes at baseline for students in the treatment and control groups. We do not find differences in six out of the seven variables we test. The only statistical difference we find is that the share of girls among treatment students (52%) is higher than in the control group (43%). Other variables do not present statistically significant differences. It is worth noting that in all the regressions we control for baseline test scores. We also present regressions in which we show that controlling for gender does not change the results in a meaningful way.

In terms of attrition, at baseline the main sample includes 2,926 students attending the schools that participate in this study. We were able to collect information for 2,400 students in the first follow-up in December 2012, which implies an attrition rate of 17.9 percent. The main reasons for not finding a student at the first follow-up was that the student had withdrawn from the school or he was absent from school the day the instruments were applied. The attrition rate is not statistically different between treatment and control groups[18]. Panel B of Table 5 presents a re-analysis of balance for the follow-up sample of December 2012. Results are very similar to those for the baseline (Panel A): only the gender variable is unbalanced. This implies that baseline characteristics of attriting and non-attriting students are not significantly different between treatment groups. In the second follow-up (December 2013), we were able to collect information for 2,446 students, which implies an attrition rate of 16.4%. Panel C of Table 5 presents an analysis of balance for this survey and, as before, we find a similar pattern to those of Panels A and B.

c. Short-term Results

We start discussing results of the follow-up collected in December 2012, at the end of the 2012 school year. This corresponds to the end of the implementation of the Mimate program for the cohort of students for which we applied the baseline in March 2012.

First, we present impact estimates of the Mimate program on the overall EGMA test scores, considering all the dimensions, in column 1 of Table 6. The estimate is statistically significant and implies an impact of the Mimate program of 0.10σ in favor of students in the treatment group. As a benchmark, Näslund-Hadley, Parker and Hernández-Agramonte (2014) document an improvement score of 0.16σ in an inquiry-based preschool mathematics program in Paraguay. Banerjee and colleagues (2007) report that two hours per day of individual instructions with tutors improved mathematics learning by 0.14σ among Indian fourth-graders. Thus, the impact effects slightly below those of international mathematics interventions.

Next, we study the potentially different effects of the program in the numeracy and shapes abilities that the Mimate program aims to develop. This is important as our implementation analysis showed that a

[18] Results not shown here but available from the authors upon request.

higher proportion of the sessions focused on shapes were implemented compared to sessions focusing on numeracy skills. We present the results in columns 2 and 3 of Table 6. Results imply a slightly bigger impact (0.12σ) on an index that considers just the abilities related to shapes in EGMA than on an index considering just the abilities related to numeracy in EGMA (an impact of 0.10σ). These results are consistent with the fact that the sessions related to shapes had a higher degree of implementation than the sessions related to numbers. In addition, the estimate for numeracy was closer to the overall program estimate than the estimate for shapes, which may imply stronger effects for numeracy.

To formally test for the effect of the share of implemented sessions on learning outcomes, we run regressions of the share of Mimate sessions implemented in each school (this is equal to 0 for all control schools) on the overall EGMA score. This is presented in column 4 of Table 6. We find that the coefficient of the share of sessions implemented is equal to 0.15 (significant at the 1% level). However, these results assume that the actual number of sessions is not endogenous to other unobserved school characteristics. To test the last point, we present in column 5 of Table 6 IV estimates where we use both the treatment indicator and the interaction between the treatment indicator and the duration of the teacher strike in each school as instruments for the share of implemented sessions.[19] Results show an estimate of 0.14, which is statistically indistinguishable from the OLS estimate.[20] These results imply that an increase in the number of sessions from what we observed on average (about 57 sessions) to the planned number of sessions (86) would increase the effect of the program by 0.05. These results also imply that differences in the share of sessions related to numeracy and shapes (25%, accordingly to our process data) would translate into a difference of 0.035 standard deviations in terms of the effects in both dimensions, which is above what we observe in columns 2 and 3 of Table 6 (0.02 standard deviations). Thus, we interpret these results as suggesting that the number of implemented sessions may explain the slightly larger program effect on skills related to shapes.

Next, we present estimates for the different items included in the EGMA test. We do this to learn which dimensions of mathematics learning are affected the most by the Mimate program. We present both naïve (without considering the potential existence of a multiple hypothesis testing problem) and corrected

[19] In the first stage we find that the coefficient of the treatment status is 1.01 (with a standard error of 0.05) and for the interaction is -0.0041 (with a standard error of 0.006). This implies that 20 additional days of the teacher strike decrease the share of sessions by 8.2%. Notice also that the average duration of the strike was 49 days for treatment schools.

[20] We also run an additional exercise in which we control for all the school variables included in Table 3 as correlates of the share of implemented sessions. Results in both the first- and second-stages do not change in a significant way (e.g., the second stage estimate is now 0.138).

(considering multiple hypothesis testing) p-values in Table 7. First, in terms of numeracy skills, we find statistically significant effects for number selection (effect of 0.18σ, significant at the 1% level), additive composition (effect of 0.15σ, significant at the 5% level), and oral counting (effect of 0.11σ, significant at the 5% level). Second, in terms of abilities related to shapes, we find a significant effect for geometric shapes (with an effect of 0.20σ that is significant at the 1% level).

We now present in Table 8 estimates to test for heterogonous treatment effects.[21] First, we study whether Mimate has a differential effect by gender. This is important because: (i) there are gender gaps in learning outcomes at baseline (see Table 5), (ii) gender was not balanced across treatment and control students at baseline, (iii) several papers document large differences in terms of mathematics teaching across genders, with a bias for boys, and (iv) the results of the teacher survey suggest that the program affected significantly the beliefs of teachers regarding a gender unbiased approach. Our results show no statistically significant treatment differences between boys and girls. This implies, on the negative side, that Mimate was unable to close the existing gender gap but, on the positive side, that it is a gender-neutral program.[22] Mimate inherently addresses the gender bias generally found when teaching Mathematics by requiring regular attention to each student, regardless of sex, and thus preventing the monopoly of attention that fidgety and high-energy boys can sometimes demand from their teachers at the expense of better-behaved girls. In fact, as we discussed before, (i) the qualitative findings indicate that in all schools using Mimate, boys and girls were included in activities equally, (ii) results in Table 3 suggest that the number of implemented sessions was similar between classes with more girls and classes with more boys, and (iii) the answers to the teacher survey strongly suggest a more gender-neutral approach to teaching Mathematics among treated teachers.

Second, we study whether there are heterogeneous effects by the language of the student. This is a dimension that poses challenges, given the difficulty of adapting the materials to different languages. Moreover, we found gaps at the baseline. Results in column 2 of Table 8 show that there are no statistical differences between students who speak Quechua or are bilingual and Spanish speakers. If anything, the treatment effects seem to be smaller for Spanish-speaking students. Thus, as in the case of gender, Mimate

[21] We only present heterogeneous effects for the overall EGMA test to save space and because we do not find systematic differences in the pattern of heterogeneous effects when analyzing effects for the numeracy and shapes dimensions of the test. All heterogeneity analyses presented here are based in the original evaluation plan.
[22] The existence of boys-biased programs in schools is not unfamiliar to Peru. Results from a science program showed that the program widened the achievement gap between boys and girls (Beuermann et al., 2013).

seems to be a program that is neutral in terms of the student's home language. This is very remarkable as our implementation analysis found that schools with more Spanish-speaking students implemented more sessions of the program.

Next, we study whether results are different for students from high and low SES households (column 3 of Table 8). As previously discussed, the exposure to different levels of mathematical reasoning for students from different SES families seems to have strong effects in terms of mathematics learning. Yet the Mimate program did not favor the higher SES sectors. Results imply that there are no statistical differences and that, if anything, children from low SES improved at a faster rate than children from high SES. The same result emerges when comparing students from rural and urban areas. These results echo the findings of Ramani and Siegler (2011), providing another example that fun, interactive mathematical games are also beneficial for low SES students.

We also study whether the effects of the Mimate program vary by different school characteristics. First, we consider potential differences between schools located in urban and rural areas (column 4 in Table 8). We do not find statistical differences among students attending urban and rural schools and, if anything, the effects for students attending rural schools are larger than for students attending urban schools. Again, this is a striking result given the above documented challenges faced by rural schools. We then consider the effect of class size (column 5 in Table 8). Teachers face the difficulty of keeping a large classroom focused on learning, but by emphasizing collaborative work in small groups, Mimate appeared to have neutralized the effect of class size as well. Students in large classrooms improved equally, on average, compared to students in smaller classrooms. Finally, we study whether the teacher's educational level generates differential program effects. Results in column 6 of Table 8 suggest that this is the case. Teachers with university degrees have stronger effects than teachers with a tertiary degree. In fact, the effect of the Mimate treatment is 0.15σ for schools in which the treatment was implemented by teachers with university degrees in contrast to an effect of just 0.04σ (and not statistically different from 0) for schools in which the program was implemented by teachers with a tertiary, non-university degree. The inherent classroom difficulties in Peru, along with a completely new pedagogical model, can explain why teachers who have university-level teaching degrees saw their students improve significantly more with Mimate in the short-term than teachers with a tertiary, non-university degree.[23]

[23] It is interesting to notice that we also find that years of teaching experience did not have a significant impact on the effects of the program.

In all, the results in Table 8 suggest that the Mimate program did not have heterogeneous effects for students with different characteristics, a positive finding given the status quo of big gaps in several dimensions, such as gender, language, income, and geographic location. The only difference we found was related to the fact that Mimate program seems to have a larger impact when the teachers have more years of education.

Next, we present the results of treatment effects using quantile regressions. This exercise allows us to estimate the impact of the treatment on the distribution of test scores (Table 9). In mathematics, teachers often find it especially difficult to find a balanced lesson plan that stimulates the curiosity of the high-achieving students without confusing the low-achieving ones.[24] Mimate's scaffolding model, which adapts to meet the individual needs of students, may help to mitigate this problem, as is suggested by the class observations and the answers to the teacher survey. Results in Table 9 imply that the treatment effects are significantly stronger for students in the bottom part of the test score distribution than for students in the top part of the distribution. For instance, while the impact for students in the bottom 10% and 25% of scores improved by between 0.12σ and 0.16σ (depending on the test considered), treatment effects for students in the top 10% and 25% of the score distribution were just between 0.04σ and 0.08σ. This shows an equitable impact of the program, and it suggests that Mimate may even narrow the achievement gap between the highest- and lowest-achieving students.

Finally, Table 10 presents treatment effects of the Mimate program on non-mathematics dimensions of the learning process. We focus on the Raven cognitive development test and on a test of early literacy skills to study whether the program may have positive or negative externalities. Results show that the program did not have statistically significant effects on these dimensions. This suggests that neither the additional learning of mathematics skills created positive externalities on other areas nor that the new focus on mathematics skills and teaching distracted resources from learning in other dimensions.

In sum, the short-term results imply that the Mimate program had a statistically significant effect on general mathematics learning outcomes, with slightly stronger effects on abilities related to understanding shapes. In terms of heterogeneous effects, the program did not produce differences against girls, students from low socioeconomic status, students whose home language is different from Spanish, or students who

[24] There is a general literature on the effects of several innovations to try to deal with class heterogeneity, such as tracking (see Duflo et al., 2011 and the references therein) and the use of software that allows students to advance at a different pace (see Barrow et al., 2009 and the references therein).

live in rural areas. Also, we do not find differentiated effects for classrooms of different sizes, but we do find stronger effects when teachers with university degrees implemented the program—in comparison to teachers with tertiary level degrees.

d. Medium-term Results

We now move to the discussion of the results of the follow-up implemented in December 2013 at the end of the 2013 school year. This corresponds to a follow-up after the students attended school for one year after the program ended and *without* receiving the program and allows us to study the medium-term effects of the intervention. We present results for the same estimations used in the short-term follow-up.

In terms of the effects of Mimate on the overall EGMA test, we do not find statistically significant effects (Table 6, Panel B). This implies that the effect of the program that we found in the short-term does not persist in the medium-term. This result is similar to what has been previously found in the literature on the effects of mathematics pre-school programs in primary education (see Duncan and Magnuson, 2013). However, as we discussed above, given the large difference in terms of program implementation in the areas related to shapes and numeracy and the consequent differences we find in their EGMA scores, we also expect to find a differential effect of the Mimate program in these areas. Results in Panel B of Table 6 confirm this, as we find that Mimate has an effect that is not statistically significant for the numeracy score but a statistically significant impact for geometric shapes with an effect of 0.15σ (significant at the 5% level using the FWER p-values). This result is interesting as we find that the strongest program effect in the short-term is also the one that has an effect in the medium-run. It is important to stress that we are very conservative in reaching this conclusion, as we are correcting our p-values for multiple hypotheses testing and, therefore, it is highly unlikely that this result is an outcome obtained by chance.

In terms of heterogeneous effects of the intervention, we do not find many significant effects, with the exception again of a positive and statistically significant interaction between the Mimate treatment and teachers with university degrees. This interaction has almost the same effect in both follow-ups (0.11σ in the short-term follow-up and 0.12σ in the medium-term) and might suggest a potential complementarity between Mimate and the educational level of the teachers. In fact, this result is consistent with the argument in Yoshikawa et al. (2013) that "structural quality" is a necessary condition for "process quality" in preschool education.

For students located at different points in the outcome distribution (Panel B of Table 9), we find that the significant effects for shapes are concentrated among the students located in the top of the distribution. This result suggests the existence of a specific type of complementarity in which after the program ended higher-achieving students (benefit more in the medium-term than lower-achieving students. In the short-run, (when the program is being implemented) the opposite is true. More research is needed to have a clear explanation for this finding.

6. Discussion and Conclusions

There is no silver bullet curriculum that solves all education problems, but tested pedagogical strategies backed by theory are the best hope for delivering returns to investments in education (Clements and Sarama, 2011; Duncan and Magnuson, 2013). The importance of mathematics skills at the preschool level has become increasingly apparent with an emerging number of randomized controlled trials using vast samples in North America and Europe. In Peru, where underachievement in mathematics persists from preschool all the way into secondary schools, a well-designed program could replicate or exceed North American and European effect sizes. In theory, Mimate provides children with age-appropriate materials that can bolster their foundations in mathematics and trains teachers to become listeners and facilitators of each child's development.

The results in this paper suggest that Mimate has statistically significant impacts after one-year of implementation and some persistent effects one year later. The effects seem to be slightly stronger in outcomes related to forms and shapes than to numeracy in the short-term and the effects are only persistent in the medium term for items related to shapes. This is consistent with the process information we present in this paper, which suggests that the implementation of the program was higher exactly in sessions related to this area. We do not find evidence that the implementation was significantly different for schools and teachers with different characteristics, suggesting that the variation in the implementation was closely related to the interruption of the program due to teacher strikes. Thus, our results suggest that the heterogeneous implementation of the program is probably driving the different effects in the numeracy and the shapes outcomes. It remains to be seen whether a complete implementation of the program can improve the impacts of the numeracy items both in the short- and in the medium-term.

In terms of cost-effectiveness, the complete program has an annual average cost of about $37 per student per year (on top of the regular expenditures). Considering that just about two-thirds of the program was implemented and given the impacts we find, we estimate that $100 increases about 0.4 standard deviations. If we compare this to some of the results reported by Kremer et al. (2013) for innovation for primary education in developing countries, we find that the program is less cost-effective than the most efficient innovations these authors review, but that it is more cost-effective than most of the interventions involving teachers.

There are several implications and questions for future research and experimentation that can be derived from our results. Dramatic changes, like asking experienced teachers to step down from the pulpit and provide inquiry-based and tailored instruction based on each student's learning needs, can take some time to take hold. The results that better qualified, but not more experienced, teachers taught the program more effectively suggest that teaching flexibility, or tools of teaching learned later in school, are critical to the program's success. In this dimension, one could think of teacher training visits or directed special attention to teachers with lower levels of education as interventions to help close this gap. However, further research would be needed to establish this connection.

Another dimension that may be relevant is extending Mimate into the home with simple take-home games (e.g. dice, puzzles). This could increase the intensity of the treatment, especially in the case of teacher strikes. Parents would then have tangible resources to engage in mathematics and mathematics talk with their children.

References:

Algan, Y, P. Cahuc, and A. Shleifer (2013). "Teaching Practices and Social Capital," American Economic Journal: Applied Economics, American Economic Association, vol. 5(3), pages 189-210, July

Almond, Douglas, and Janet Currie. (2010) “Human Capital Development Before Age Five,” Handbook of Labor Economics, 4B, 1316–1476.

Anderson, M. (2008) ““Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects”. Journal of the American Statistical Association 103(484): pp. 1481-1495.

APOYO (2012) “Propuesta de implementación de un programa educativo para fomentar capacidades matemáticas en niños y niñas de 5 años en aulas de Ayacucho y Huancavelica” Mimeo.

Araujo, C., P. Carneiro, Y. Cruz, and N. Schady (2016) “Teacher Quality and Learning Outcomes in Kindergarten”, Quarterly Journal of Economics, 131(3).

Aunola, et. al. (2006) “Developmental dynamics between mathematical performance, task motivation, and teachers’ goals during the transition to primary school”. British Journal of Educational Psychology, 76, 21–40.

Banerjee, A., C. Shawn, E. Duflo, L. Linden. 2007. “Remedying Education: Evidence from Two Randomized Experiments in India.” The Quarterly Journal of Economics, 122(3) 1235-1264.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2016) “Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of “Teaching at the Right Level” in India” NBER Working Paper No. 22746.

Banerjee, A. and E. Duflo (2010) Poor Economics. Public Affairs.

Barro, R. and J-W. Lee (2013) Barro-Lee Educational Attainment Dataset http://barrolee.com

Barrow, L., L. Markman, and C. Rouse, (2009), “Technology's Edge: The Educational Benefits of Computer-Aided Instruction,” The American Economic Journal: Economic Policy, (1), 52-74.

Berlinski, S. and N. Schady (2015) *The Early Years Child Well-Being and the Role of Public Policy*. Interamerican Development Bank.

Beuermann, D., E. Naslund-Hadley, I. Ruprah, and J. Thompson (2013). "The Pedagogy of Science and Environment: Experimental Evidence from Peru," Journal of Development Studies, vol. 49(5), pages 719-736.

Bos, M.S., A. Elías, E. Vegas, and P. Zoido (2016). Latin America and the Caribbean in PISA 2015: How did the region perform? Brief #1. Inter-American Development Bank.

Centro Peruano de Audición, Lenguaje y Aprendizaje (CPAL). 2012. “Informe sobre análisis del instrumento: Evaluación Experimental del Programa MIMATE-Inicial en los Departamentos de Ayucucho y Huancavelica, Perú,” Unpublished IDB Consultancy Report.

Clements D.H., Sarama J. (2008) Experimental evaluation of the effects of a research-based preschool mathematics curriculum. American Education Research Journal; 45(2):443–494.

Clements, D. H., Samara, J. (2009). Early Childhood Mathematics Education Research: Learning trajectories for young children. New York, NY: Routledge.

Clements, D. H., Samara, J. (2011). “Early Childhood Mathematics Intervention.” Science 333(6045): 968 –70.

Cruz-Aguayo, Y., J. LoCasale-Crouch, S. Schodt, T. Guanziroli, M. Kraft-Sayre, C. Melo, S. Hasbrouck, B. Hamre, and R. Pianta. 2015. “Early Schooling Classroom Experiences in Latin America: Focusing on What Matters for Children’s Learning and Development.” Mimeo, IADB.

Cunha, F. and J. Heckman (2007) "The Technology of Skill Formation." American Economic Review, 97(2): 31-47.

de Grave, W. S., Dolmans, D. H. J. M., & van der Vleuten, C. P. M. (1999). Profiles of effective tutors in problem-based learning: Scaffolding student learning. Medical Education, 33, 901-906.

Duflo, E., R. Glennerster, M. Kremer (2008). “Using Randomization in Development Economics Research: A Toolkit” In T. Schultz and John Strauss, Eds., Handbook of Development Economics. Vol. 4.

Duflo, E., P. Dupas, and M. Kremer (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." American Economic Review, 101(5): 1739-74.

Duncan, G. et. al. “School Readiness and Later Achievement”. Developmental Psychology 2007, Vol. 43, No. 6, 1428–1446

Duncan, G. and K. Magnuson. (2013). "Investing in Preschool Programs." Journal of Economic Perspectives, 27(2): 109-32.

Ellis, E.S., Worthington, L. (2004). Research Synthesis on Effective Teaching Principles and the Design of Quality Tools for Educators Technical Report No. 5 produced for the National Center to Improve the Tools of Educators, University of Oregon

Englert, C. S., Raphael, T. E., Anderson, L. M., Anthony, H. M., & Stevens, D. D. (1991). Making strategies and self-talk visible: Writing instruction in regular and special education classrooms. American Educational Research Journal, 2, 337-372.

ESCALE. Estadística de calidad educativa. Ministerio de Educación, Perú. http://escale.minedu.gob.pe/

Geary, D.C., M. Hoard, L. Nugent, D. Bailey (2013). “Adolescents’ Functional Numeracy Is Predicted by Their School Entry Number System Knowledge”. PLOS ONE 8 (1).

Gersten R, Jordan N.C., Flojo J.R. (2005). Early identification and interventions for students with mathematics difficulties. Journal of Learning Disabilities;38(4):293–304. [PubMed: 16122059]

Gertler, P et al. (2014) “Labor Market Returns to Early Childhood Stimulation: A 20-year Follow-up to an Experimental Intervention in Jamaica", Science, 2014, Volume 344, Issue 6187.

Hammil, D. D., N. A Pearson, J. K. Voress, y M. Frostig. 1995. DTVP: Método De Evaluación De La Percepción Visual De Frostig. Español: 2a ed. MP, 45, 1-4. México, D.F.: Editorial El Manual Moderno.

Havnes, T. and M. Mogstad (2011). "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes," American Economic Journal: Economic Policy, vol. 3(2), pages 97-129, May.

He, F., L. Linden, and M. MacLeod (2009) "A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trail." Mimeo, University of Texas, Austin.

Heckman, J. and S. Mosso (2014). "The Economics of Human Development and Social Mobility," Annual Review of Economics, Annual Reviews, vol. 6(1), pages 689-733.

Hogan, K., & Pressley, M. (1997). Scaffolding scientific competencies within classroom communities of inquiry. In K. Hogan & M. Pressley (Eds.), Scaffolding student learning: Instructional approaches & issues (pp. 74-107). Cambridge, MA: Brookline.

Innovation for Poverty Action (IPA). 2012. “Mimate Inicial Evaluación Experimental del programa de Matemáticas para Inicial en los Departamentos de Ayacucho y Huancavelica, Perú,” Unpublished IDB Consultant Report.

Inter-American Development Bank (2018). Centro de Información para la Mejora de los Aprendizajes. https://www.iadb.org/es/cima

Jordan, N. Huttenlocher, J., and S. Levine. (1992). Differential calculation abilities in young children from middle and low-income families. Developmental Psychology; 28:644–653.

Jordan, N. Locuniak, M. (2009) Early Math Matters: Kindergarten Number Competence and Later Mathematics Outcomes. *Dev Psychol*. May; 45(3): 850–867.

Kim JS, Burkhauser MA, Quinn DM, Guryan J, Kingston HC, and Aleman K. (2017) “Effectiveness of structured teacher adaptations to an evidence-based summer literacy program.” *Reading Research Quarterly*. Forthcoming.

Kremer, M., C. Brannen, and R. Glennerster (2013) “The Challenge of Education and Learning in the Developing World” Science, Vol. 340 no. 6130 pp. 297-300, April.

Lepper, M. R., Drake, M. F., & O'Donnell-Johnson, T. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds.), Scaffolding student learning: Instructional approaches & issues (pp. 108-144). Cambridge, MA: Brookline.

Liddell, C. and G. Rae (2001) “Predicting early grade retention: A longitudinal investigation of primary school progress in a sample of rural South African children” British Journal of Educational Psychology 71 (3): 413–428,

Lynch K, Kim JS (2017). Effects of a summer mathematics intervention for low-income children. Educational Evaluation and Policy Analysis. Forthcoming.

Malofeeva E, Day J, Saco X, Young L, Ciancio D. (2004). Construction and evaluation of a number sense test with Head Start children. Journal of Educational Psychology; 96(4):648–659.

Muralidharan, M., A Singh, and AJ Ganimian (2017) “Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India” NBER Working Paper No. 22923

Naslund-Hadley, E., S. Parker, J. Hernández-Agramonte (2014) “Fostering Early Math Comprehension: Experimental Evidence from Paraguay”. Global Education Review 1(4): 135-154.

National Mathematics Advisory Panel. Foundations for success: The final report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education; 2008.

Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R. M., & Morrison, F. J. (2008a). Classroom effects on children’s achievement trajectories in elementary school. *American Educational Research Journal, 45(2),* 365–397.

Pianta, R. C., K. M. La Paro, and B. K. Hamre. (2008b). *Classroom Assessment Scoring System (CLASS) Manual: Pre-K*. Baltimore, MD: Brookes Publishing Co.

Paxson, C. and N. Schady, 2010. "Does Money Matter? The Effects of Cash Transfers on Child Development in Rural Ecuador," Economic Development and Cultural Change, vol. 59(1), pages 187-229, October.

Ramani, G. B. & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. Journal of Applied Developmental Psychology, 32, 146-159.

Romano E., Babchishin L., Pagani L.S., Kohen D. (2010). School readiness and later achievement: replication and extension using a nationwide Canadian survey. Dev Psychol Sep;46(5): 995-1007.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia supported learning environments. Educational Technology Research and Development, 50(3), 77-96.

Schady, N., J. Behrman, M. Araujo, R. Azuero, R. Bernal, D. Bravo, F. Lopez-Boo, K. Macours, D. Marshall, C. Paxson and R. Vakis (2015) “Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries.” Journal of Human Resources 50 (2).

Seo, K.-H., & Ginsburg, H. P. (2004). What is developmentally appropriate in early childhood mathematics education? Lessons from new research. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), Engaging young children in mathematics: Standards for early childhood mathematics education (pp. 91-104). Hillsdale, NJ: Erlbaum

Stith, A., Gorman, K.S & Choudhury, N. (2003). Beyond primary school: Psychosocial risk and gender. Applied Psychology: An International Review. 52(4), 614-629.

Vygotsky, Lev. (1978) Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Walker, S. P., S. M. Chang, C. A. Powell, and S. M. Grantham-McGregor (2005). “Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study.” The Lancet 366 (9499), 1804–1807.

Westfall, P. H. & Young, S. S. (1993). Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. John Wiley and Sons.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L., Gormley, W. T., Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M (2013) “Investing in Our Future: The Evidence Base on Preschool,” Society for Research in Child Development, Washington, D.C. (2017).

Appendix 1: Exercises included in the "Early grade Mathematical assessment" (EGMA)

The quantitative tests were preschool adapted versions of the "Early Grade Mathematical Assessment" (EGMA) originally developed by the Research Triangle Institute International. The test measures various abilities related to mathematics with the following exercises:

a) *Comparing quantity:* Children are tested for their understanding of "more," "less," and "equal" with an image exercise comparing rows of kittens, chickens, and bunnies. The tester challenges the child to indicate in each row which box had more, less, or equal numbers of animals.
b) Basic shape recognition: Children are presented with a circle, triangle, square, and rectangle and asked to correctly name each shape.
c) Advanced shape recognition: The same shapes are dispersed in a matrix of shapes, and the children are asked to match four plastic tiles (a circle, square, triangle, and rectangle) with their corresponding shapes.
d) *Basic object counting:* Children are asked to point with their finger and count balloons in a picture of balloons numbered 1, 2, 3, 4.
e) *Advanced object counting:* The same exercise is repeated with a picture of twelve balloons numbered 1 through 12.
f) *Number selection:* Children view a grid of twelve boxes, each one containing a different number of stars rising from 1 to 12. Then, they are asked to point out which box has three stars, six stars, nine stars, and twelve stars.
g) *Advanced number selection:* Children view three boxes containing clusters of four hearts, five hearts, and eleven hearts respectively. Below the boxes are the numbers 5, 11, and 4 placed out of order. The children are asked to match each symbolic number with the appropriate box of hearts.
h) *Fine motor skills:* Children copy to the best of their ability images of basic shapes, symbolic numbers and letters. Scores were later calculated by a team of digitation specialists.
i) *Symmetry:* Children are shown an image of a butterfly and asked to draw a line over the butterfly that divides it into two equal parts.
j) *Advanced symmetry:* Children try to match one side of a house with one of three options to complete the picture.

k) *Shape sequence:* Unfinished patterns of shapes (triangle, square, circle, triangle, square…etc.) are presented to the children. Children indicate which shapes on the right side of the page would complete the pattern for each row. Patterns with color were also tested and color blindness did not affect performance.

l) *Number sequence (clock):* Confronted with an image of a clock face, children respond to the questions "Which number comes after 4?" and "Which number comes before 9?"

m) *Number sequence (calendar):* The tester shows children a calendar page of April and asks the children to help reschedule a party by answering the question "What day is two days after April 5th?"

n) *Additive composition:* On the left side of the page is a box displaying three kittens and on the right side are pairs of similar boxes showing another numbers of kittens. Children are asked to identify the correct pair of boxes (one kitten and two kittens) that together are the same as the box on the left. The exercise is repeated three more times with flowers, apples, and hearts.

o) *Geometric shapes:* Children are given four plastic triangle tiles and a plastic rhombus tile and asked to arrange them to cover up a large hexagon shape depicted on the page. Then, the tester takes away two triangles and gives the child a rhombus and asks them to complete the task again.

p) *Addition and subtraction word problems*: "Daniel has one dog. María has one dog. How many dogs do they have in total?" and "There are four children walking to school. Two of them are boys and the rest are girls. How many girls are walking to school?"

Table 1
Global Raven cognitive test scores

Country	Score	Standard deviation	Average Age	Source
Peru	12.3	5.5	5.2	This study
United Kingdom	15.1	3.7	5.3	Gathercole, 1994
Denmark	18.5	2.4	5	Olsen 1992
India	12	4.5	5	Bhogle 1992
Poland	15.1	4.5	5	Szustrowa 1992
Slovakia	14.4	3.8	5.7	Ferjencik 1992
Argentina	14	3	5	Leibovich de Figueroa 1992
Canada	14.2	-	5.3	Wright 1995
Global Average	14.5	3.8	5.1	

Notes: Raven test comparison between countries.

Table 2

Process Information: Planned and Implemented Sessions

Module (by date)	Number of sessions	Average of sessions completed (%)
	March	
Basic shapes	2	100.00%
Spatial relationships	7	100.00%
	April	
Bend and cut	4	100.00%
Cover figures I	3	100.00%
	May	
Construct with cubes	3	100.00%
Number sequence I	3	100.00%
Dice images	4	100.00%
	June	
Define quantities I	6	100.00%
Assemble quantities	2	100.00%
Order numbers I	4	100.00%
	July	
Quantity representations	3	100.00%
	August	
Draw figures and patterns	3	99.39%
Symmetry	5	95.27%
	September	
Cover figures II	2	83.64%
Pre-writing I	2	69.09%
Recognize shapes and space	4	50.91%
Pre-writing II	1	43.64%
Cognitive game	2	39.09%
	October	
Pre-writing III	1	32.73%
Bend	5	17.45%
Number sequence II	2	14.55%
Define quantities II	2	6.36%
Pre-writing IV	1	5.45%
	November	
Order numbers II	3	0.00%
Pre-writing V	1	0.00%
Number patterns	7	0.00%
	December	
Numbers for measuring	2	0.00%
Number representations	2	0.00%
Total number of sessions	86	

Total avg. % of completed sessions	65.96%

Notes: Comparison between planned and actually completed number of sessions by module. The order of the modules in this table follows the MIMATE timeline. Data comes from the Mimate program.

Table 3
Determinants of the Percentage of Completed Sessions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baseline overall Math score	0.087							-0.032
	(0.042)							(0.056)
% women per (treated) school		0.062						0.141
		(0.077)						(0.086)
% of quechua speaking households per (treated) school			-0.067					-0.031
			(0.029)					(0.045)
% of bilingual speaking households per (treated) school			-0.081					-0.083
			(0.036)					(0.051)
% of high income families per (treated) school				0.067				0.006
				(0.034)				(0.050)
Urban schools					0.057			0.044
					(0.022)			(0.040)
Average class size per school						0.003		-0.000
						(0.002)		(0.003)
% of teachers with university degree per (treated) school							0.020	0.038
							(0.026)	(0.027)
Observations	43	43	43	43	43	43	40	40
R-squared	0.121	0.012	0.142	0.070	0.122	0.096	0.014	0.225

Notes: Dependent variable is the percentage of completed sessions per (treated) school, which is an average of the percentage of completed sessions of all the (treated) classes in each school. Each column shows the correlation between the dependent variable and some potential determinants. Robust standard errors in parentheses.

Table 4
Educational outcomes by groups (standardized effects)

	Boys	Girls	Rural	Urban	Spanish	Quechua	Bilingual
Mathematics	0.02	-0.04	-0.37	0.09	0.15	-0.46	-0.31
Cognitive ability	0.04	-0.05	-0.28	0.08	0.15	-0.36	-0.28
Writing	0.01	-0.01	-0.59	0.16	0.22	-0.64	-0.42
Oral comprehension	0	-0.01	-0.4	0.11	0.17	-0.44	-0.36
N	1529	1397	623	2303	1809	393	243

Notes: Baseline average deviation from the sample mean of each subgroup for different outcomes. Standardized effects for the complete sample. Data comes from the Mimate program.

Table 5

Sample Balance

	Treatment	Control	Difference	t-stats
Panel A: Baseline survey				
Score on baseline math	-0.04	0.03	-0.07	-0.83
Score on baseline cognitive	-0.08	0.08	-0.16	-1.31
Score on baseline early literacy	-0.07	0.08	-0.15	-1.24
Female	0.52	0.43	0.09	3.00
Attended preschool age 3 and 4	0.86	0.84	0.02	0.66
Speaks Quechua at home	0.19	0.18	0.01	0.08
Attends a rural school	0.21	0.21	0.00	-0.04
Panel B: Short-term follow-up				
Score on baseline math	-0.04	0.01	-0.05	-0.49
Score on baseline cognitive	-0.10	0.01	-0.11	-0.88
Score on baseline early literacy	-0.08	0.02	-0.10	-0.75
Female	0.51	0.44	0.07	2.33
Attended preschool age 3 and 4	0.87	0.82	0.05	1.21
Speaks Quechua at home	0.19	0.20	-0.01	-0.21
Attends a rural school	0.21	0.24	-0.03	-0.38
Panel C: Mid-term follow-up				
Score on baseline math	-0.03	0.04	-0.07	-0.71
Score on baseline cognitive	-0.07	0.07	-0.14	-1.14
Score on baseline early literacy	-0.05	0.09	-0.14	-1.14
Female	0.51	0.44	0.07	2.46
Attended preschool age 3 and 4	0.87	0.83	0.04	1.04
Speaks Quechua at home	0.19	0.19	0.00	0.00
Attends a rural school	0.21	0.21	0.00	-0.08

Notes: Data comes from the Mimate program.

Table 6
Effects on Mathematics Outcomes

Panel A: Short-term effect

Dependent Variable	Overall Test	Shapes Items	Numeracy Items	Overall Test	Overall Test
	(1)	(2)	(3)	(4)	(5)
Treatment	0.10	0.12	0.10		
	(0.03)	(0.04)	(0.03)		
Share of Completed Sessions				0.15	0.14
				(0.05)	(0.05)
Controls	Yes	Yes	Yes	Yes	Yes
Estimation Method	OLS	OLS	OLS	OLS	IV
Observations	2400	2400	2400	2152	2152

Panel B: Medium-term effect

	Overall Test	Shapes Items	Numerac y Items	Overall Test	Overall Test
	(1)	(2)	(3)		
Treatment	0.00	0.06	-0.03		
	(0.03)	(0.03)	(0.05)		
Share of Completed Sessions				0.00	0.00
				(0.02)	(0.02)
Controls	Yes	Yes	Yes	Yes	Yes
Estimation Method	OLS	OLS	OLS	OLS	IV
Observations	2416	2416	2416	1895	1895

Notes: Standard errors corrected for heteroskedasticity and intra-cluster correlation at school level in parenthesis. Data comes from the Mimate program.

Table 7

Effects by items of the EGMA test

Panel A: Short-term effect				Panel B: Medium-term effect			
Dependent variable	Coefs.	Naïve p-value	FWER p-value	Dependent variable	Coefs.	Naïve p-value	FWER p-value
Understanding shapes							
Geometric shapes	0.195	0.000	0.000	Geometric shapes	0.159	0.002	0.030
	(0.0371)				(0.0522)		
Shape recognition	0.143	0.016	0.127	Shape recognition	0.0691	0.267	0.516
	(0.0592)				(0.0623)		
Spatial ability	0.130	0.022	0.147	Spatial ability	-0.0773	0.192	0.516
	(0.0566)				(0.0592)		
Symmetry	0.0985	0.113	0.397	Symmetry	0.0933	0.059	0.309
	(0.0621)				(0.0494)		
Figure sequence	0.0688	0.257	0.527				
	(0.0607)						
Reproduce figures	0.0669	0.384	0.527	Reproduce complex figures	0.0663	0.185	0.516
	(0.0768)				(0.0500)		
Numerical Literacy							
Number selection	0.178	0.000	0.011	Number selection	0.00914	0.893	1.000
	(0.0489)				(0.0678)		
Additive composition	0.148	0.002	0.036	Additive decomposition	-0.0246	0.694	0.999
	(0.0476)				(0.0625)		
Oral counting	0.105	0.002	0.036	Oral counting	-0.018	0.644	0.999
	(0.0340)				(0.039)		
Addition and subtraction word problems	0.118	0.012	0.112	Addition and subtraction problems	0.000527	0.992	1.000
	(0.0469)				(0.0560)		
Comparing quantity	0.0962	0.066	0.350	Comparing quantity	-0.0849	0.187	0.865
	(0.0523)				(0.0643)		
Naming numbers	0.0939	0.085	0.358	Naming numbers	-0.0724	0.267	0.916
	(0.0545)				(0.0652)		
Advanced numeration	0.0818	0.135	0.432	Advanced numeration	0.0151	0.749	0.999
	(0.0547)				(0.0473)		
Number sequence	0.0704	0.159	0.435	Ordering numbers	-0.0501	0.442	0.986
	(0.0500)				(0.0651)		
Comparing numbers	0.0525	0.224	0.448	Comparing numbers	-0.0613	0.315	0.945
	(0.0431)				(0.0610)		
Basic numeration	0.0202	0.660	0.681	Number sequence	0.0298	0.662	0.999
	(0.0461)				(0.0682)		
				Previous and subsequent	-0.00556	0.934	1.000
					(0.0672)		
				Measurement units	-0.0701	0.216	0.880

	(0.0567)		
Writing numbers	-0.0323	0.651	0.999

Notes: Naive p-values correspond to standard p-values of the individual regressions. FWER p-values correct for the familywise error rate (the probability of making a type I error) and use a Westfall-Young (1993) type correction as explained by Anderson (2008). Data comes from Mimate program.

Table 8
Heterogeneous effects: Overall index

Subgroup(s):	Female	1-Quechua 2-Both	High socio-economic level	Urban	Section size	Teacher w/ university degree
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Short-term effect						
	Mathematics					
Treatment	0.11	0.08	0.11	0.15	0.13	0.04
	(0.04)	(0.04)	(0.04)	(0.07)	(0.10)	(0.05)
Subgroup 1 x Treatment	-0.00	0.02	-0.06	-0.06	-0.00	0.11
	(0.03)	(0.09)	(0.06)	(0.08)	(0.00)	(0.06)
Subgroup 2 x Treatment		0.07				
		(0.07)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2400	2037	1149	2400	2400	1901
Panel B: Medium-term effect						
	Mathematics					
Treatment	-0.01	0.00	0.00	0.01	0.01	-0.05
	(0.04)	(0.04)	(0.04)	(0.08)	(0.10)	(0.05)
Subgroup 1 x Treatment	0.00	-0.01	0.04	-0.00	-0.00	0.12
	(0.04)	(0.10)	(0.05)	(0.09)	(0.00)	(0.06)
Subgroup 2 x Treatment		0.01				
		(0.10)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2416	2202	1091	2416	2293	1811

Notes: All columns control for baseline score, school locality, test giver. Only column (2) considers two subgroups and two interactions: Quechua speaking household and household that speak Quechua and Spanish equally. Section size in column (5) is a continuous variable. Standard errors corrected for heteroskedasticity and intra-cluster correlation at school level in parenthesis. Data comes from the Mimate program.

Table 9
Effects according to baseline mathematics quantiles

Panel A: Short-term effect

	Overall	*Shapes*	*Numbers*
	(1)	(2)	(3)
q10 Treatment	0.12	0.15	0.16
	(0.04)	(0.05)	(0.05)
q25 Treatment	0.14	0.15	0.13
	(0.03)	(0.03)	(0.04)
q50 Treatment	0.09	0.13	0.09
	(0.02)	(0.02)	(0.02)
q75 Treatment	0.08	0.08	0.05
	(0.03)	(0.02)	(0.02)
q90 Treatment	0.04	0.05	0.05
	(0.02)	(0.02)	(0.02)
Controls	Yes	Yes	Yes
Observations	2400	2400	2400

Panel B: Medium-term effect

	Overall	*Shapes*	*Numbers*
	(1)	(2)	(3)
q10 Treatment	-0.04	0.02	-0.04
	(0.04)	(0.03)	(0.04)
q25 Treatment	-0.01	0.02	-0.03
	(0.03)	(0.03)	(0.03)
q50 Treatment	-0.02	0.03	-0.01
	(0.03)	(0.02)	(0.02)
q75 Treatment	0.03	0.07	0.00
	(0.02)	(0.03)	(0.02)
q90 Treatment	0.04	0.10	-0.01
	(0.02)	(0.11)	(0.02)
Controls	Yes	Yes	Yes
Observations	2416	2416	2416

Notes: Standard errors corrected for heteroskedasticity and intra-cluster correlation at school level in parenthesis. Data comes from the Mimate program.

Table 10

Effects on Other Learning Outcomes

	Cognitive abilities	Early literacy
	(1)	(2)
Panel A: Short-term effect		
Treatment	0.06	-0.02
	(0.08)	(0.05)
Controls	Yes	Yes
Observations	2387	2365
Panel B: Medium-term effect		
Treatment	0.030	0.031
	(0.059)	(0.060)
Controls	Yes	Yes
Observations	2,111	2,068

Notes: Standard errors corrected for heteroskedasticity and intra-cluster correlation at school level in parenthesis. Data comes from the Mimate program.