

Changing Pedagogy to Improve Skills in Preschools: Experimental Evidence from Peru

Francisco A. Gallego, Emma Näslund-Hadley, and Mariana Alfonso

Abstract

Changing pedagogical practices is a promising, cost-effective avenue for improving education in developing countries, especially when done without changing current inputs such as teachers and instruction time. This article presents the results of a randomized evaluation of a program that aimed at changing the pedagogical approach used to teach the existing national mathematics curriculum. The program provides tools to regular preschool teachers to use an inquiry- and problem-based learning approach to tailor instruction to preschoolers in Peru. The results show an improvement of overall mathematics outcomes, which persist for some content areas even one year after the program ended. In contrast to results from previous research that suggest mathematics programs are biased along gender and socioeconomic lines, there is no evidence of differential effects by gender, language spoken at home, or proxies for socioeconomic status. Results also imply persistent stronger impacts on students whose teachers have university degrees.

JEL classification: I21, I28, O15

Keywords: education, pedagogical innovations, tailored instruction, mathematics, early childhood development

1. Introduction

There is wide agreement on the importance of improving skills for the benefit of subsequent economic and social performance, especially among poor children. However, it is less clear which specific pedagogical

Francisco A. Gallego (corresponding author) is a professor at Pontificia Universidad Católica de Chile and affiliated professor at J-PAL, Santiago, Chile; his email address is fgallego@uc.cl. Emma Näslund-Hadley is a lead education specialist in the Education Division at the Inter-American Development Bank; her email address is emman@iadb.org. Mariana Alfonso is a lead sector specialist at the Climate Change and Sustainable Development Department at the Inter-American Development Bank; her email address is marianaa@iadb.org. The research for this article was financed by the Inter-American Development Bank, J-PAL, and the Government of Japan through the IDB JPO Trust Fund. The authors thank Eric Edmonds (the editor), three anonymous referees, Samuel Berlinski, Jessica Goldberg, Lorenzo Neri, Elizabeth Spelke, and participants in seminars at Harvard, PUC-Chile, the 2018 RIDGE Impact Evaluation Forum, and the Workshop on “Equity: Contributions from Research in Education, Economy, Psychology and Neurosciences” held at PUC-Chile in March 2016 for insightful comments. Thanks also go to Inés Levy, Sergio de Marco, Cristine von Dessauer, César Huaroto, and Alejandro Saenz for expert research assistance, and to Kathryn McLellan for editorial help. The program studied in this paper (Mimate) was designed and implemented by Instituto Apoyo in coordination with the Ministry of Education of Peru (MINEDU) and the Inter-American Development Bank (IDB). Innovations for Poverty Action (IPA) assisted with the sampling, design field work, and statistical analysis of the data. 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. A supplementary online appendix is available with this article at *The World Bank Economic Review* website.

approaches and techniques can improve learning for students at different educational levels. This is especially relevant for preschool programs both because of the relevance of early childhood development and preschool education for subsequent development of cognitive and noncognitive skills (Cunha and Heckman 2007) and because of the wide heterogeneity in terms of effectiveness of different preschool programs (e.g., Duncan and Magnuson 2013; Dillon et al. 2017).

This paper explores the effects of the Mimate program, which provides tools to *regular* preschool teachers to help teach the *existing* national math curriculum using a scaffolding approach based on inquiry- and problem-based learning during normal school hours.¹ The program seeks to change the traditional vertical teaching model of memorization and repetition—where lessons are heavy on teacher discourse, all students work on the same activities, and there are few student-led exploration activities—into a horizontal model (APOYO 2012). The program is organized around 86 sessions, of 45 minutes each, during regular school hours. The implementation of the program includes the use of educational materials, organizational tools, and teacher training and support. The annual cost per student of scaling up the program is \$37.

This article tries to answer two main research questions in this context: first, whether it is possible to change how teachers teach by providing them with a pedagogical model, including tools and training on how implement the model; and second, whether a more student-centered pedagogy is cost-effective. Answering both questions requires several changes in the way education is conducted, which is challenging in a context of preschoolers in a developing country, with several classroom difficulties such as large class sizes, limited educational materials, and poor infrastructure.²

This paper presents the results of an impact evaluation of this program using a randomized controlled trial (RCT) implemented in schools located in and around three Peruvian cities: Huancavelica, Angaraes, and Ayacucho, using a stratified randomization with 54 treatment and 53 control schools, stratifying by cities and urban/rural area. The results imply that, at the end of the program, the intervention increased mathematics outcomes for children who attended the treated schools by between 0.14 and 0.21 standard deviations (σ hereafter) of the control group for the overall math test. In terms of specific math content areas, the estimates imply treatment effects of between 0.11σ and 0.19σ for items related to numeracy and by between 0.18σ and 0.23σ for those related to understanding shapes. Next, the article reports treatment effects on the same outcomes for the same students one year *after the program ended* (called “one-year follow-up” in the paper), and results imply that the effects decrease in magnitude. The treatment effects for the overall mathematics test and for the numeracy section are not different from 0, and treatment effects for the shapes section range between 0.06σ to 0.13σ . 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; Dillon et al. 2017).³ As discussed below, the larger impacts for understanding shapes may be a consequence of the intensity of implementation.

- 1 Scaffolding is an adaptive interactive strategy in which 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). This is closely related to recent research on the effects of the “teaching at the right level” approach, which also seeks to focus the instruction on each student’s needs (Banerjee et al. 2016).
- 2 This article extends to preschool education a recent literature related to the development of cost-effective pedagogical innovations that change the way teachers teach, holding expenditure, staffing, and instructional time constant (e.g., Barrow, Markman, and Rouse 2009; Kremer, Brannen, and Glennerster 2013; Jackson and Makarin 2017; Muralidharan 2017). In turn, regarding the second question, the analysis is closely related to recent research on the effects of the “teaching at the right level” approach, which also seeks to focus the instruction on each student’s needs (Banerjee et al. 2016).
- 3 However, in this literature, results often fade out to then show up later again in the form of better labor market outcomes (Chetty et al. 2011). There is no information available to test for this long-run effect in the case of Mimate.

This article also studies the possible existence of heterogeneous effects along several observable dimensions related to characteristics of both students and schools by estimating interaction effects with characteristics such as the student's gender, language, and socioeconomic status; school location; class size; and teachers' education. Results imply that the treatment does not have statistically and persistent effects based on these dimensions except for teacher education. Results indicate that in both follow-ups, when the teacher who implemented the program has a university degree,⁴ the program has a larger and statistically significant effect. This suggests that there is a complementarity between the teacher's human capital and the program, which probably indicates that pedagogical innovations like Mimate demand a certain minimum level of teacher human capital to work successfully.

The article then presents estimates of quantile regressions to understand heterogeneous effects across the distribution of outcomes. The results imply that the program has significant effects for shape scores in the first follow-up for students in all quantiles (with impacts between 0.10σ and 0.35σ) but with stronger effects for students in the lower quantiles. The impacts decrease in magnitude in the one-year follow-up and are statistically significant only among the students located in the top of the distribution (for students located in quantiles 75 to 95). The change in the gradient of the effects across quantiles in both follow-ups seems to be driven by a decrease in the program's impact on students in the lower quantiles, as the effects for students in the top quantiles are not statistically different across follow-ups. Quantile regression results for numeracy scores resemble estimates for shape scores in the first follow-up, but are mostly flat and close to 0 in the one-year follow-up. These results are important in understanding the mechanisms behind the effects of programs like Mimate and can be consistent with a combination of static and dynamic complementarities of the program with each student's ability.

To facilitate the interpretation of the results, the article also uses administrative data, class observations, and surveys given to teachers and parents to analyze the program's actual implementation. The process evaluation suggests that, on average, 66 percent of the 86 planned lessons were implemented. This partial implementation was mainly due to a national teacher strike that canceled up to three months of class time depending on the school. This may explain the findings, given that sessions on number sequence, quantities, and patterns came at the end of the program and were covered less intensively. In contrast, the sessions on shapes and figures were covered at the beginning of the school year. In fact, the process evaluation reported in this article finds that while 82 percent of the sessions planned to cover figures were implemented (equivalent to about 30 sessions), just 57 percent of the sessions related to numbers were implemented (equivalent to about 25 sessions).⁵

In addition, results from class observations and a survey given to teachers after the program suggest that the program affected students' and teachers' classroom behavior, teachers' beliefs about their students, and the teachers' ability to meet the objectives and to teach mathematics (with respect to the control group). Differences in treatment effects among teachers with and without a university degree make it possible to understand the mechanisms behind the program's impact. The results also imply that the differences in favor of treated teachers related to improvements in the perceptions of student behavior and abilities are stronger for those with university degrees. The estimates also imply that while treated teachers with university degrees are more likely to state that they have enough time to cover the materials and have a more positive attitude toward the teaching of math compared to their similarly educated peers in the control group, the same result does not hold for teachers without university degrees. In contrast, results for teaching support, availability of materials, and other dimensions related to the program do not differ based on teacher education.

4 In Peru, teachers can have either 1) a tertiary, nonuniversity degree in teaching granted by a teacher-training institute; 2) a degree in teaching granted by a university (and advanced degrees such as MAs or PhDs).

5 This is a very coarse classification between numbers-related and shapes-related sessions, as the program seems to influence both dimensions in an interlocking way. However, the study wants to illustrate the differences between the dimensions based on the number of actual lessons taught.

Although there is a large literature exploring the causal impacts of preschool programs (see the review in Almond and Currie 2010) and exploring the effects of pedagogical innovations (see the review in Kremer, Brannen, and Glennerster 2013), few studies have conducted randomized controlled trials to study the effects of implementing pedagogical innovations among preschoolers in the developing world (He, Linden, and MacLeod 2009; Naslund-Hadley, Parker, and Hernández-Agramonte 2014; Dillon et al. 2017).⁶ Thus this paper makes several contributions to the existing literature. First, this study represents one of the few randomized evaluations of mathematics programs that combine a relatively large sample of schools and students and follows up the students after the program ended to identify fade-out effects.⁷ Second, this paper contributes to the research on the effects of 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.⁸ Third, this paper contributes to the literature on the impact of classroom-based interventions that try to change pedagogical practices without affecting the current stock of teachers (He, Linden, and MacLeod 2009; Banerjee and Duflo 2010; Barrow, Markman, and Rouse 2009; Muralidharan, Singh, and Ganimian 2018). Fourth, the finding that there is a complementarity between pedagogical innovations and teacher human capital is new in the literature on developing countries.⁹

The remainder of the paper is structured as follows. Section 2 describes the context, the program, the experimental design, and the data used. Section 3 presents the main results. Section 4 discusses mechanisms and cost effectiveness, and section 5 briefly concludes.

2. Intervention and Study Design

Context: Peruvian Education

Access to education in Peru is high, with net school enrollment rates in primary and secondary education above 95 percent and 75 percent respectively in 2013.¹⁰ Although learning outcomes have also improved considerably in the past decade, they are still low compared to other countries and present steep socioeconomic gradients. At 15 years of age, according to the 2015 PISA report, Peruvian students ranked 62nd in mathematics, 63rd in reading, and 64th in science out of the 70 countries that participated. In addition, Peru 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 second-grade national census evaluation show a

6 This article does 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, which shows that children brought up in a more favorable early environments 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 Dillon et al. (2017) test a mathematics preschool program based on games in small groups of children in India, but there are two key differences with respect to this paper. This article analyzes an intervention implemented in schools by regular teachers in regular classrooms with about 30 students per class and focuses on an intervention based on tailoring instruction at the student level.

8 This is related to both the “teaching at the right level” (Banerjee and Duflo 2010; Banerjee et al. 2016) and the scaffolding (Heckman and Mosso 2014) empirical literatures. 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), this study presents a randomized evaluation of a large-scale intervention using the scaffolding approach for preschoolers in a developing country.

9 While this result contrasts with Jackson and Makarin (2017), who show that a low-cost intervention produces stronger effects for less-educated teachers in the United States, note that there is a big difference in the level of human capital of teachers in both studies. While in the sample of this article 44 percent of the teachers have a tertiary, nonuniversity degree, 44 percent have a university degree, and just 12 percent have a postgraduate degree; in their sample 40 percent of the teachers have MA or PhD degrees, and all the others have university-level degrees. Thus, the sample of less-educated teachers in that paper is equivalent to the more educated teachers in the sample of this article.

10 See IADB (2018), accessible at <https://www.iadb.org/es/cima>.

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 between Peru's urban and rural areas.

The Peruvian Ministry of Education (MINEDU) has recognized preprimary school as a priority for improving educational outcomes. As a result, from 2007 to 2011 MINEDU increased preprimary education spending per student by 70 percent, compared to increases of 60 percent and 46 percent for primary and secondary education respectively (ESCALE 2019). In parallel, the rate of enrollment in preschools for children aged three to five years increased from 53 percent to 75 percent in the 2001–2012 period, with increases for children living in both rural and urban areas (enrollment increased from 53 percent to 75 percent in urban areas and from 44 percent to 66 percent 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 percent to 12 percent between 2000 and 2013.

This effort to expand access to preprimary education, however, was not initially accompanied by programs aimed at changing the pedagogical practices in order to improve learning outcomes. Thus, the status quo in terms of pedagogical practices in preprimary education was one of memorization and repetition using a model of vertical pedagogy where all the students in the classroom received the same materials and tended to repeat, at the same time, the teacher's instructions—for instance, all students would sing at the same time or repeat a number after the instructor (APOYO 2012).

The Mimate Program: Design and Implementation

In this context, MINEDU opted to implement a pilot of the Mimate program in a sample of schools as a pedagogical innovation. The salient characteristic of Mimate is the use of an inquiry-based approach. Inquiry-based learning, or co-operative learning as Vygotsky (1978) called it, is based on the idea that social interaction is essential for learning. Fundamentally, it seeks to flip the traditional Peruvian vertical teaching model of memorization and repetition into a horizontal, student-centered model in which students progress at their own pace. This is closely related to recent research on the effects of the “teaching at the right level” approach, which also seeks to focus the instruction on each student's needs (Banerjee et al. 2016).¹¹ The conceptual motivation for this type of intervention is related to finding efficient ways of solving the problem of the organization of teaching in a context in which there are heterogeneous students, multitasking, economies of scale, and several other binding constraints (e.g., Barrow, Markman, and Rouse 2009; Kremer, Brannen, and Glennerster 2013; Jackson and Makarin 2017; Muralidharan 2017).

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.” Scaffolding seeks to develop activities that are tailored to the individual child's ability level so they are neither too hard nor too easy. Therefore, Mimate includes both hard and soft scaffolding techniques combined with a workbook, twice-a-month formative assessments, and visits from teacher assistants who help teachers learn how to provide good-quality interactions.¹²

- 11 This is also related to the strand of 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).
- 12 Soft scaffolding is human support provided by a teacher or peers that helps students meaningfully participate in their learning. 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 for each student (de Grave, Dolmans, and van der Vleuten 1999; Hogan and Pressley 1997; Lepper, Drake, and O'Donnell-Johnson 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 consists of three 45-minute sessions per week, fitting easily into the daily schedule of Peruvian preschools. The teacher begins the class with a quick overview of the day's objectives, then quickly splits up the children into small groups or pairs for activities, and then the class finishes with a short group-discussion and review of the materials covered. The curriculum and lesson plan advance with numerical challenges that gradually progress from very basic to advanced. More than simply increasing in difficulty, each task prepares the student to tackle the next one. For example, kids write numbers first as dots (•, ••, etc.) to prepare them for writing down symbolic numbers. By the end of the curriculum, students should be manipulating symbolic numbers fluidly and pointing out number patterns in their daily lives.¹³ Mimate follows the national curriculum of Peruvian preschools. Therefore, the program, like the national curriculum, is organized around the topics of numerical literacy and understanding shapes.¹⁴ Table 1 presents the planned sequence of sessions by topic and month.

Hands-on teaching materials are critical to the mathematics lessons, and each student receives a personal box of materials that includes a gamebook, 13 pennies, stickers, 2 red and blue cardboard circles, 24 numbered pictures of bugs and butterflies, 8 small wooden blocks, 1 small mirror, plastic tiles of rhombuses, 1 regular dice, 1 dice with shapes on each side, 1 dice with colors on each side, 12 crayons, and a plastic ball among other things.

In addition, the program handed out the following materials for the teachers: an instruction book (see below), a book of game instructions, a large dice (20 centimeters per side), a CD with music, a clock, a calendar, an organizer, and a set of big stickers with numbers from 1 to 12. In turn, the program also provided the following classroom materials: three sets of cards and the materials for seven different board games. Those items should be kept in the "Mimate corner" of the classroom throughout the school year so that children can play with these toys and tools during their free time.

Individualized instruction requires teachers to know exactly what material each child understands. This is no simple task in classrooms of up to 30 students per teacher.¹⁵ Mimate developed a solution that allows for accurate formative assessments using a simple five-minute round of flash cards between the teacher and each individual student. 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.

A key component of the program is the teacher role. One important challenge for the program was large content and pedagogical gaps among teachers. Therefore, the program developed several strategies

- 13 This is one example of the different approaches between the Mimate program and regular preschools (the counterfactual situation). In a class on numeral recognition, a control school-teacher writes numerals on the blackboard and asks the children first to say them out loud and then to practice writing them. However, although a student may know that "4" is called *four*, he or she may not comprehend what "4" means. Instead in Mimate classrooms, students write numbers first as dots to prepare them for writing symbolic numbers. In one activity to support the transition to numerals, pairs of students play bingo by rolling a dice and placing markers on the corresponding numeral on their respective boards.
- 14 Numerical literacy includes both (i) number sequence and number aspects (in which 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) and (ii) capturing structure (in which children practice the rule of cardinality, that the last number spoken when counting signifies the total number. When counting is understood, children are taught to consider numbers in small groups and learn how to combine and separate groups mentally without relying on counting). In turn, understanding shape includes (i) variety of geometric shapes (in which 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) and (ii) training fine motor skills (in which 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 physical learning along with language. The activities hone the coordination of fingers while reinforcing the lessons of earlier modules).
- 15 Recently, sophisticated computer programs have become popular tools to test students and recommend further exercises based on their specific capabilities (Barrow, Markman, and Rouse 2009; Muralidharan, Singh, and Ganimian 2018). However, these types of programs were not feasible to implement in the Peruvian context as only 7 percent of schools had Internet access and computer skills among teachers were sparse.

Table 1. Process Information: Planned and Implemented Sessions

Module (by date)	Number of sessions	Average of sessions completed (%)
<i>March</i>		
Basic shapes	2	100.00
Spatial relationships	7	100.00
<i>April</i>		
Bend and cut	4	100.00
Cover figures I	3	100.00
<i>May</i>		
Construct with cubes	3	100.00
Number sequence I	3	100.00
Dice images	4	100.00
<i>June</i>		
Define quantities I	6	100.00
Assemble quantities	2	100.00
Order numbers I	4	100.00
<i>July</i>		
Quantity representations	3	100.00
<i>August</i>		
Draw figures and patterns	3	99.39
Symmetry	5	95.27
<i>September</i>		
Cover figures II	2	83.64
Prewriting I	2	69.09
Recognize shapes and space	4	50.91
Prewriting II	1	43.64
Cognitive game	2	39.09
<i>October</i>		
Prewriting III	1	32.73
Bend	5	17.45
Number sequence II	2	14.55
Define quantities II	2	6.36
Prewriting IV	1	5.45
<i>November</i>		
Order numbers II	3	0.00
Prewriting V	1	0.00
Number patterns	7	0.00
<i>December</i>		
Numbers for measuring	2	0.00
Number representations	2	0.00
Total number of sessions	86	
Total avg. % of completed sessions	65.96%	

Source: Data for this table come from the administrative dataset collected by the Mimate program. The order of the modules in this table follows the Mimate timeline.
Note: The table reports the number of planned sessions and the share of sessions actually implemented by topic and month. The table also reports the total number of planned sessions and the percentage of completed sessions at the bottom of the table.

to provide teachers with hands-on support to ensure consistent lesson delivery across classrooms. This was developed mainly through three main sets of tools:

1. *Instruction book.* Each teacher received a 123-page instruction book with the following information: (i) a set of general instructions, including good teaching practices and the four areas of emphasis as mentioned above; (ii) instructions for use of the materials related to the program and the organization of the classroom; (iii) suggestions for other activities to be developed, including a suggested meeting

with parents at the beginning of the program; (iv) detailed instructions for each session including scripted step-by-step instructions and examples of phrases and conversations; and (iv) a short review of the psychological and mathematical foundations of the program. This book was meant to act as a handbook for implementation.

2. *Preprogram teacher training.* Teachers participated in a three-day training session in which they learned the program's objectives and were trained in the program's sessions and materials.
3. *Teacher support during the program's implementation.* Teacher coaches observed sessions and gave advice on how to improve teaching based on Mimate's dictates. These support sessions took place during the school year (from March to November). The maximum number of planned visits was nine. The coaches were typically retired teachers who received special training. The main idea of the support was to ensure the quality of teacher-student interactions. In each visit the coach observed a class and, then met with the teacher one on one to give feedback and support.

Information to characterize the actual implementation of the Mimate program was collected (reported in [table 1](#)). The program planned for 86 sessions in total to be implemented from March to December. Analyses using administrative data imply that, on average, 66 percent of the program's sessions were taught. However, this varies by topic: While 82 percent of the sessions directly related to shapes were implemented, only 57 percent of the sessions directly related to numeracy were covered. This is a consequence of timing. Numeracy topics were planned to be covered more intensively in the second part of the year when a teacher strike canceled up to three months of classes in some schools.¹⁶ This article also studies whether there are systematic differences across schools in terms of the program's implementation (see table S2.1 in the supplementary online appendix). While some school-level variables are statistically significant when explaining the number of sessions implemented, the effects are not economically relevant, and thus there is not first-order heterogeneity in the number of sessions implemented.

To assess the quality of the program's implementation, class observations were collected from a random sample of 37 classrooms in treated schools. Results imply that all treatment schools have the "Mimate corner" in the classroom and that all teachers were using the Mimate materials as suggested by the program. In turn, in 78 percent of the schools the actual set-up of the Mimate corner was consistent with the design suggested by the program, and the board games provided by Mimate were available to students in 92 percent of the schools. These results suggest that implementation fidelity was very high and that teachers were using Mimate as designed.

In terms of attendance at the teacher training and support related to its pedagogical approach, 83 percent of the teachers attended the three sessions before the program started, and teachers received on average six visits of in-class training.¹⁷ In terms of individual determinants of the share of training sessions completed by teachers in treatment schools, findings show that the only variable that is statistically

16 In total, 87 percent of teachers in treatment schools participated in the strike at some point in time. When running a regression of the number of sessions implemented on the number of days the teachers participated in the strike, the estimated coefficient is -0.0030 (with a standard error of 0.0007). This implies that 20 additional days of the teacher strike decreased the share of sessions by 6 p.p. Notice also that the average duration of the strike was 84 days for treatment schools. There are no comparable data for teachers in control schools. However, using a survey of school principals, there are no statistical differences across treatment and control schools in terms of a proxy for the intensity of support for the strikes at the school level (a question on whether there was full support for the strikes in the school): 87 percent in treatment schools and 82 percent in control schools, with the difference not being statistically significant with a p -value of 0.49 .

17 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. Sixty-three percent of the teachers attended these sessions.

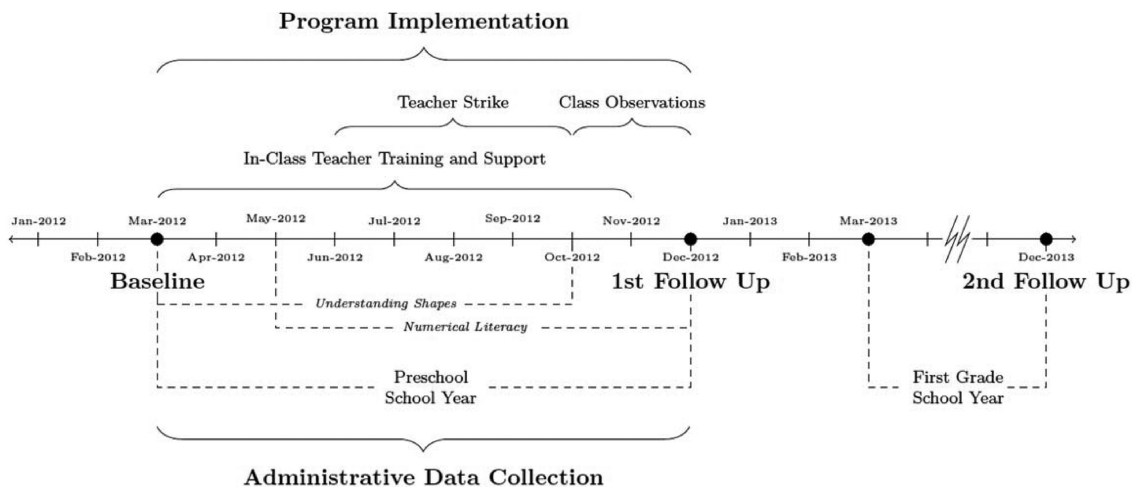
significant is whether the school is located in an urban area, with schools receiving about one training session less than the teachers in rural schools (see table S2.1).¹⁸

Research Design and Data

A randomized controlled trial allocating treatment at the school level is the basis of the study design. The sample of this evaluation was determined by the Ministry of Education of Peru and corresponds to preschool centers with at least six students enrolled in the five-year-old grade¹⁹ in 2011 in the cities of Huancavelica, Angaraes, and Ayacucho. This produces a sample of 107 schools, which were randomly assigned into 53 treatment schools and 54 control schools using a stratified randomization with six strata, defined using the combination of the school location (urban/rural) and the city in which the school is located. The treatment group would adopt the Mimate program in their five-year-old classrooms (kindergarten), while the control group would continue to use traditional pedagogical practices (as discussed above, this generally means teaching practices based on frontal instruction, memorization, and repetition).

Figure 1 presents the timeline associated with the activities of both implementation and evaluation of the Mimate program. 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 “first follow-up”). A second follow-up was collected 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 “one-year follow-up”). 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 out of a total baseline sample of 2,926 students.

Figure 1. Program and Evaluation Timeline



Source: Figure constructed by the authors using the timeline of the intervention and the data-collection process.
Note: This figure presents a timeline of the main stages related to the intervention and the data collection process.

- 18 The program also included a suggested initial meeting with parents at the beginning of the program, where teachers could explain the program and its objectives. Nearly all—96 percent—of the classrooms did so with an average attendance of 17.3 parents (the average number of students by classroom is 23.1). There are no detailed data available regarding the contents of this session, as this was a suggested session and the objectives were very general.
- 19 In Peru, preschool or initial education covers ages three to five. In graded preschools, each age group is in a different grade.

The general format of these data collection processes was very similar across them, consisting of student tests at the baseline and two follow-ups, and parent and teacher surveys at the baseline and first follow-up. To test for the development of mathematics skills, the analysis is done using preschool-adapted versions of the “Early Grade Mathematical Assessment” (EGMA) originally developed by the Research Triangle Institute International. Appendix S2 in the supplementary online appendix provides a detailed description of the items covered in EGMA. This article presents results of the items included in EGMA at three different aggregation levels: (1) an overall index that includes all the items in EGMA; (2) two indices by curricular area, one for items related to numerical abilities and the other for items related to shapes; and (3) measures of the development of each ability included in EGMA.²⁰ In addition to mathematics tests, the analysis also uses information from instruments to assess the development of nonmathematics outcomes: the Raven test of general cognitive ability and a test of early literacy skills.²¹

The questionnaires given to teachers and parents collected information about the child’s classroom and home experience. The information from these surveys serves two purposes. The first is to study treatment heterogeneous effects. Second, the information from these surveys makes it possible to understand the mechanisms through which the program may have affected learning outcomes.

Lastly, between the months of October and December of 2012 an external team visited 44 randomly selected classrooms (in 37 treatment and 7 control schools) to verify the intensity of treatment (i.e., teacher and student attendance records and the level of implementation of different Mimate lessons). In addition to observing teacher-student interactions and student-student interactions, a sample of 12 classrooms in treatment and 4 classrooms in control schools were randomly selected to be videotaped during mathematics lessons,²² which were analyzed using the CLASS (Class Assessment Scoring System) rubric.²³

Sample, Balance, and Attrition

Table 2 compares outcomes at the baseline for variables collected for students, teachers, and parents in the treatment and control groups. This also makes it possible to characterize the sample used in this paper. Forty-eight percent of the students are girls, they are five years old on average, 90 percent of them had enrolled in preschool at age 3 or 4, 19 percent speak Quechua at home, 22 percent of them attend a rural school, and 72 percent of them receive help at home with their homework.²⁴ In terms of data from the baseline parental survey, 77 percent of the kids live in a household where the father is present, 21 percent

20 The preschool version of EGMA was originally developed for Paraguay and was validated for alignment with the Peruvian 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, Pearson, and Voress (1995).

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

22 Financial restrictions did not make it possible to expand the sample to more schools for the class observation or to videotape more classes.

23 These class observations were only used by the evaluation team (and not used in any form to give advice to teachers or schools in terms of teaching practices). See Pianta, La Paro, and Hamre (2008) for a description of the CLASS instrument, and Araujo et al. (2016) and Cruz-Aguayo et al. (2015) for applications in Latin American countries. CLASS 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. For cost reasons, the analysis does not include information from other instruments (for example, the TIMSS video study) that would have made it possible to measure the precise pedagogical practices used by teachers.

24 An SES index was also constructed, which is described in table S2.2 of the supplementary online appendix. This index corresponds to the sum of four dummy variables indicating: (1) whether the family owns a cell phone, (2) whether the family owns books, (3) whether the family owns a TV, and (4) whether the family owns a computer. The average for the sample is 2.93.

Table 2. Balance Test

	Control (1)	Treatment (2)	Difference (3)
<i>Panel A: Student sample</i>			
Score on baseline math [#]	0.00 (1.00)	-0.11 (1.01)	-0.11 [0.10]
Score on baseline cognitive tests [#]	0.00 (1.00)	-0.14 (0.91)	-0.14 [0.11]
Score on baseline writing test [#]	0.00 (1.00)	-0.17 (0.97)	-0.16 [0.09]
Female	0.44 (0.50)	0.52 (0.50)	0.08*** [0.03]
Attended preschool age 3 and 4	0.90 (0.30)	0.89 (0.31)	0.01 [0.02]
Speaks Quechua at home	0.18 (0.39)	0.19 (0.39)	0.01 [0.08]
Attends a rural school	0.22 (0.41)	0.21 (0.41)	-0.01 [0.08]
Age	5.03 (1.10)	5.00 (1.38)	-0.03 [0.07]
SES (index)	2.94 (0.99)	2.92 (1.00)	-0.02 [0.13]
Has help at home for homework	0.71 (0.45)	0.73 (0.44)	0.02 [0.04]
<i>Panel B: Teacher sample</i>			
Age	38.6 (7.2)	41.0 (8.4)	2.4 [1.6]
Female	1.00 (0.00)	1.00 (0.00)	0.00 [-]
University degree	0.58 (0.50)	0.61 (0.49)	0.03 [0.10]
Tenured	0.57 (0.50)	0.63 (0.49)	0.06 [0.11]
Union member	0.28 (0.46)	0.22 (0.42)	-0.06 [0.11]
Number of students in the classroom	20.34 (6.32)	22.75 (7.52)	2.40 [1.52]
Teaches mostly using Quechua	0.23 (0.42)	0.16 (0.37)	-0.07 [0.09]
Teaches using both Quechua and Spanish	0.35 (0.48)	0.44 (0.50)	-0.09 [0.11]
<i>Panel C: Parent sample</i>			
Father lives with child	0.77 (0.41)	0.75 (0.43)	-0.02 [0.03]
Mother's education (high school graduate)	0.21 (0.41)	0.21 (0.41)	-0.00 [0.04]
Father unemployed	0.11 (0.32)	0.11 (0.31)	-0.00 [0.02]
Has educational books at home	0.72 (0.45)	0.70 (0.46)	-0.02 [0.05]
Parenting (index)	4.02 (1.77)	3.95 (1.70)	-0.07 [0.16]
Child expected education (index)	4.05 (0.91)	4.07 (0.86)	0.02 [0.15]
Family income (in soles)	641.3 (805.0)	665.0 (874.4)	23.7 [104.3]

Source: Data used in this table come from the baseline information collected using instruments applied to students, a survey applied to teachers, and a survey applied to parents of the students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: Columns (1) and (2) present estimated averages and standard deviations for subjects in treatment and control groups. Column (3) present estimates of the differences and standard errors in square brackets. [#]denotes a variable standardized to the control group. See table S2.2 in the supplementary online appendix for definitions of the indices used in this table. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

of their mothers are high school graduates, 11 percent of their fathers are unemployed, 72 percent of the students have educational books at home, and monthly family income is 654 soles (about US\$245 at the market exchange rate). The average parent expects his or her child to receive university education, and they report spending four days per week in several parenting activities (see the definition of the parenting and child expectation indexes in table S2.2 of the supplementary online appendix). All of the teachers are female; they are 40 years old on average; 59 percent of them have university degrees; 23 percent of them teach mostly using Quechua and 35 percent using both Quechua and Spanish; 25 percent of them are union members; and 60 percent of them are tenured. Their class size is 21.6 students on average.

In terms of balance, there are no statistical differences in 24 out of the 25 variables in [table 2](#). The only statistical difference found is that the share of girls among treatment students (52 percent) is higher than in the control group (43 percent). Other variables do not present statistically significant differences; F tests of the joint significance of all the variables to predict treatment status in each panel imply that it is not possible to reject the null hypotheses that the variables at baseline for each group are not correlated with treatment status (with p -values of 0.20, 0.45, 0.68 for panels A, B, and C, respectively).²⁵

In terms of attrition, the attrition rate was 17.9 percent in the first follow-up survey and 16.4 percent in the one-year follow-up survey. The main reasons for not finding a student at the follow-up measurements were that the student had withdrawn from the school or she or he was absent from school on the day when data were collected. The attrition rate in both follow-ups is not statistically different between treatment and control groups.²⁶

Econometric Models and Statistical Tests

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

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

where Y represents an outcome for student i who attends school j , T is a dummy variable that takes the 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. Then, estimates of α quantify the differences in means between students attending schools that received the Mimate program in the 2012 school year and those that did not receive it. Standard errors are clustered at the school level, and all specifications include strata-randomization dummies and dummies for the different enumerators.²⁷

The analysis considers three different specifications for the following variables included in vector X : (1) no student control variables, (2) the value of Y for student i at baseline, and (3) all the variables included in panel A of [table 2](#). If the randomization is successful, adding control variables should

25 However, the variables related to students' outcomes in the baseline (math scores, cognitive ability, and writing tests) present worse outcomes for students in the treatment group. While these are not statistically significant, their sizes are relatively large (with differences of between 0.11σ and 0.17σ). Thus, the regressions for the main exercises of the paper control for variables included in panel A of [table 2](#).

26 Table S2.3 in the supplementary on appendix presents regressions for the determinants of attrition in each follow-up survey, including treatment status and other variables. The only statistically significant effect is related to the variable measuring whether the student receives help with homework, with a positive effect in the first follow-up survey. [Table 4](#) presents estimates of treatment effects using inverse probability weighting. Treatment effects are not different from those that do not control for attrition, which is expected given the results on the determinants of attrition. In addition, table S2.4 in the supplementary online appendix presents a balance analysis for students included in both follow-up surveys, and the main results presented in [table 3](#) do not change in a significant way.

27 When analyzing results for the different components of the tests, results may be affected by an inference problem related to multiple hypotheses testing. In this case adjusted p -values are calculated following the methodology suggested by [List, Shaikh, and Yu \(2019\)](#).

not change the estimate of α but should just provide more precise estimates.²⁸ However, as presented in [table 2](#), students' gender presented statistical differences between treatment and control groups. In addition, some differences were not statistically different but had relatively large sizes (in favor of the control group). Therefore, adding them may also make it possible to correct for initial differences. In general, results imply that there are no large differences in the estimates of α when using different estimates of X in equation (1), but estimates are larger (as expected, given the initial differences in favor of the control group) and more precisely estimated.

The basic analysis is extended in two dimensions. First, the article also includes robustness exercises to deal with the effects from attrition presenting results using inverse probability weighting and Lee bounds ([Lee 1999](#)). As discussed below, this is not likely to be a concern, as attrition is not correlated with observables and the sample is balanced. Second, the article also considers heterogeneous treatment effects running regressions that include interaction effects between the treatment dummy and student/school observable characteristics of the following form:

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

where W corresponds to a variable that varies at the school or student level. Then, if the estimate of γ is statistically different from 0, it implies that the treatment effects are different for students having different W . Finally, the article also presents heterogeneous treatment effects using quantile regressions to estimate the effects of treatment across the outcomes' distribution, following a structure similar to the one presented in equation (1). In this case, the analyses present an estimator of α that is different for individuals in different quantiles of the distribution, conditional on outcomes in the baseline (as they are included in the vector of covariates jointly with all the other variables in panel A of [table 2](#)).

3. Main Results: Treatment Effects on Test Scores

Main Effects

The discussion starts with results for test scores considering the two follow-ups collected in December 2012 and December 2013. Results are presented in [table 3](#), panel A for the first follow-up and panel B for the one-year follow-up results.

First, impact estimates of the Mimate program on the overall EGMA test scores for the first follow-up are presented in columns (1) to (3) in panel A. When not controlling for baseline variables (column 1), results imply that Mimate has a statistically significant impact (at the 10 percent level) of 0.14σ in favor of students in the treatment group. In the next two columns, when adding baseline controls, the estimate increases to between 0.18σ and 0.21σ , which is expected, as the baseline test scores were higher for students in the control group. In addition, as expected, standard errors also decrease, and therefore treatment effects are now statistically significant at the 1 percent level. As a benchmark, [Araujo et al. \(2016\)](#) identified "teacher effects" (i.e., the effect of increasing the quality of the teacher by one standard deviation) of 0.09σ for a sample of kindergarten classrooms in Ecuador. Thus, this suggests that this low-cost intervention has larger impacts than important improvements in teacher quality.

Next, columns (4) to (9) present results for the potentially different effects of the program on numeracy and shape-identifying abilities. This is important, as the implementation analysis showed that a higher proportion of the shape-focused sessions were implemented. Results imply larger impacts (between 0.18σ and 0.23σ) on an index that considers just shape abilities in EGMA than on an index considering just numeracy abilities in EGMA (between 0.11σ and 0.19σ). These results are consistent with the different degree of implementation of the shapes and number sessions previously discussed.

28 For example, it is well established that tests scores at the student level present high variation. Therefore, controlling for baseline estimates reduces the idiosyncratic variation in the results and leads to more precise estimates ([Duflo, Glennerster, and Kremer 2008](#)).

Table 3. Main Treatment Effects

Dependent variable	Overall test (1)	Overall test (2)	Overall test (3)	Shapes (4)	Shapes (5)	Shapes (6)	Numbers (7)	Numbers (8)	Numbers (9)
<i>Panel A: First follow-up</i>									
Treatment	0.14* (0.08)	0.18*** (0.06)	0.21*** (0.06)	0.18** (0.09)	0.21*** (0.07)	0.23*** (0.06)	0.11 (0.08)	0.15*** (0.05)	0.19*** (0.05)
Student controls	No	BL test	All	No	BL test	All	No	BL test	All
Observations	2,400	2,400	2,276	2,400	2,400	2,276	2,400	2,400	2,276
<i>Panel B: One-year follow-up</i>									
Treatment	-0.05 (0.08)	0.01 (0.05)	0.04 (0.05)	0.06 (0.06)	0.09* (0.05)	0.13** (0.05)	-0.07 (0.08)	-0.02 (0.06)	0.01 (0.06)
Student controls	No	BL test	All	No	BL test	All	No	BL test	All
Observations	2,416	2,416	2,171	2,416	2,416	2,171	2,416	2,416	2,171

Source: Data used in this table come from the baseline information and from the first and one-year follow-up tests applied to students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: This table presents estimated treatment effects on math test scores for students in the sample. Columns (1), (4), and (7) present estimates without controls. Columns (2), (5), and (8) present estimates just controlling for baseline test scores. Columns (3), (6), and (9) control for a vector of baseline control that includes all variables presented in panel A of table 2. Robust standard errors clustered at the school level are reported in parentheses. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Panel B of [table 3](#) presents results for the one-year follow-up. In terms of the effects of Mimate on the overall EGMA test, there are no statistically significant effects (columns 1 to 3). This result is similar to what has been previously found in the literature on the effects of preschool programs on primary education (see [Duncan and Magnuson 2013](#)), and implies that the short-term effect of the program does not persist in the medium term. However, as discussed above, given the differences in terms of program implementation in the areas related to shapes and numeracy, further analyses could find a differential effect of the Mimate program in these areas. Indeed, results in panel B of [table 3](#) confirm this. Mimate has an effect that is not statistically significant for numeracy, but for geometric shapes effects are estimated in the range between 0.06σ and 0.13σ , some of which are statistically significant (columns 5 and 6).

[Table 4](#) presents two different robustness exercises to deal with the potential effects of attrition on the estimates. First, estimates of treatment effects using inverse-probability weighting, using a probit model where all the variables in panel A of [table 2](#) are used as regressors to explain attrition (the marginal effects of the probit model are presented in [table S2.3](#), columns (2) and (4), in the supplementary online appendix). Estimates of treatment effects are very similar to those that are found using an OLS approach, which is expected as attrition is mostly balanced across treatment and control groups. Second, Lee bounds are estimated, and again the results are very similar to what is in [table 3](#). For example, in the first follow-up (panel A), in all cases the lower bounds are statistically different from 0 within a range of effects between 0.15σ and 0.16σ , with upper Lee bounds reaching values between 0.23σ and 0.25σ . In the one-year follow-up (panel B), results imply larger impacts for shapes items than for numeracy items. These results are interesting because Lee bounds do not depend on observable variables to model the potential impact of attrition. To sum up, these results suggest that the potential effects of selective attrition of students do not affect the estimates in a significant way.

Table 4. Effects on Mathematics Outcomes: Robustness Checks

Dependent variable	Overall test (1)	Shapes items (2)	Numeracy items (3)
<i>Panel A: First follow-up</i>			
Treatment, IPW	0.21*** (0.07)	0.24*** (0.06)	0.19*** (0.05)
Upper and lower Lee bounds	0.16; 0.25 [0.11, 0.30]	0.16; 0.25 [0.11, 0.32]	0.15; 0.23 [0.10, 0.28]
Controls	Yes	Yes	Yes
Observations	2276	2276	2276
<i>Panel B: One-year follow-up</i>			
Treatment, IPW	0.00 (0.08)	0.12* (0.07)	−0.03 (0.05)
Upper and lower Lee bounds	−0.05; 0.12 [−0.10; 0.18]	0.05; 0.18 [−0.02; 0.25]	−0.07; 0.06 [−0.12; 0.13]
Controls	Yes	Yes	Yes
Observations	2,171	2,171	2,171

Source: Data used in this table come from the baseline information and from the first and one-year follow-up tests applied to students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: This table presents robustness checks considering potential effects of attrition. Treatment effects were estimated using Inverse-Probability-Weighting (IPW) using a probit to model attrition using all the variables included in panel A of [table 2](#) as right-hand side variables (marginal effects are presented in [table S2.3](#) of the supplementary online appendix). Lee bound estimates use the dependent variable after being partialled-out using also the variables included in panel A of [table 2](#). Standard errors were computed using block boot-strapping at the school level in parentheses for treatment effects estimated using IPW and in square brackets for Lee bounds. *** Significant at the 1 percent level. ** Significant at the 5 percent level. * Significant at the 10 percent level.

Finally, [table 5](#) presents estimates for the different items included in the EGMA test. The analyses consider both unadjusted and adjusted *p*-values (considering the potential existence of a multiple hypothesis

Table 5. Effects by Items of the EGMA Test

Panel A: first follow-up				Panel B: One-year follow-up			
Dependent variable	Coefs.	Significance level		Dependent variable	Coefs.	Significance level	
		Unadjusted	List et al.			Naïve	List et al.
<i>Understanding shapes</i>							
Geometric shapes	0.22	***	***	Geometric shapes	0.16	***	***
Shape recognition	0.15	**	**	Shape recognition	0.08		
Spatial ability	0.13	**	**	Spatial ability	−0.03		
Symmetry	0.10			Symmetry	0.08	*	
Figure sequence	0.08						
Reproduce figures	0.08			Reproduce complex figures	0.05		
<i>Numerical literacy</i>							
Number selection	0.22	***	***	Number selection	0.00		
Additive composition	0.17	***	***	Additive decomposition	−0.01		
Oral counting	0.16	***	***	Oral counting	−0.04		
Addition and subtraction word problems	0.12	**		Addition and subtraction problems	−0.02		
Comparing quantity	0.12	***	**	Comparing quantity	−0.10		
Naming numbers	0.12	**	*	Naming numbers	−0.07		
Advanced numeration	0.12	**	*	Advanced numeration	−0.05		
Number sequence	0.09	*		Ordering numbers	−0.05		
Comparing numbers	0.08			Comparing numbers	−0.02		
Basic numeration	0.03			Number sequence	0.02		
				Previous and subsequent	0.02		
				Measurement units	−0.05		
				Writing numbers	−0.02		

Source: Data used in this table come from the baseline information and from the first and one-year follow-up tests applied to students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: Each variable analyzed in this table corresponds to the residual after partialling out by the effect of variables in panel A of table 2. Treatment effect corresponds to the effect of the treatment on each variable. Unadjusted corresponds to standard significance levels of the individual regressions, List et al. is from the List, Shaikh, and Yu (2019) correction for multiple hypothesis testing. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

testing problem). First, for the first follow-up, in terms of numeracy skills, the results imply statistically significant effects for number selection (effect of 0.22σ , significant at the 1 percent level), additive composition (effect of 0.17σ , significant at the 1 percent level), oral counting (effect of 0.16σ , significant at the 1 percent level), comparing quantity (effect of 0.12σ , significant at the 5 percent level), naming numbers (effect of 0.12σ , significant at the 10 percent level), and advanced numeration (effect of 0.12σ , significant at the 10 percent level). In terms of abilities related to shapes, the results imply significant effects on geometric shapes (effect of 0.22σ , significant at the 1 percent level), shape recognition (effect of 0.15σ , significant at the 5 percent level), and spatial ability (effect of 0.13σ , significant at the 5 percent level). The fact that there are significant effects on several different dimensions suggests that the overall treatment effects are not driven by increases in just one specific dimension of mathematics knowledge. Second, regarding the one-year follow-up (panel B), results show a large and statistically significant effect only for geometric shapes of 0.16σ (significant at the 1 percent level). This result is interesting as the strongest program effect in the short term is also the one that has an effect one year after the program ended.²⁹

29 Finally, the analysis also estimates treatment effects of the Mimate program on nonmathematics dimensions of the learning process using the Raven cognitive development test and a test of early literacy skills (table S2.5 in the supplementary online appendix). Results show that the program did not have statistically significant effects on these dimensions. This suggests that the additional learning of mathematics skills did not create positive externalities on other areas, nor did the new focus on mathematics skills take resources from learning in other dimensions.

Table 6. Heterogeneous Effects: Overall Mathematics Test

Subgroup(s):	Female (1)	1-Quechua 2-Both (2)	High socio- economic level (3)	Urban (4)	Section size (5)	Teacher w/university title (6)
<i>Panel A: First follow-up</i>						
Treatment	0.23*** (0.06)	0.18*** (0.06)	0.21*** (0.05)	0.27*** (0.11)	0.25*** (0.05)	0.11 (0.09)
Subgroup 1 × treatment	−0.04 (0.05)	0.10 (0.04)	−0.02 (0.02)	−0.07 (0.13)	−0.00 (0.01)	0.18* (0.10)
Subgroup 2 × treatment		0.10 (0.10)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,276	2,121	2,276	2,276	2,112	1,792
<i>Panel B: One-year follow-up</i>						
Treatment	0.02 (0.06)	0.01 (0.06)	0.02 (0.05)	0.04 (0.13)	0.02 (0.05)	−0.08 (0.09)
Subgroup 1 × treatment	0.00 (0.06)	0.04 (0.04)	0.02 (0.04)	−0.02 (0.14)	−0.00 (0.01)	0.21** (0.11)
Subgroup 2 × treatment		0.04 (0.14)				
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,167	2,032	2,167	2,167	2,167	1,701

Source: Data used in this table come from the baseline information and from the first and one-year follow-up tests applied to students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: All columns control for all variables included in panel A of table 2. Only column (2) considers two subgroups and two interactions: Quechua-speaking households and bilingual (Quechua and Spanish) households. Section size in column (5) is a continuous variable. Standard errors corrected for heteroskedasticity and intra-cluster correlation at the school level in parentheses. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

Heterogenous Treatment Effect

This article next presents estimates to test for heterogonous treatment effects.³⁰ The discussion is concentrated on results for the first follow-up (panel A of table 6), given that there are significant short-run treatment effects.³¹ The results show no statistically significant treatment differences between boys and girls (column 1), between students who are bilingual and Spanish speakers (column 2), between students of different socioeconomic status (column 3), between students from rural and urban areas (column 4), and between students attending classrooms with different numbers of students (column 5). All these dimensions reflect specific challenges to the application of the treatment and the fact that there are no heterogeneous effects by these dimensions suggests the program was successful in being applied to different populations.

The only dimension for which results show differential program effects is the teacher’s educational level. Results in column (6) of table 6 suggest that this is the case.³² Teachers with university degrees

30 All heterogeneity analyses presented here are based on the original evaluation plan before the analyses of the first follow-up were conducted. The only heterogeneous analysis presented in the paper that were not part of the original evaluation plan are discussed in footnote 34. These analyses were added to help understand, ex-post, some of the results.

31 Table S2.6 in the supplementary online appendix presents simple correlations among the variables included in the heterogeneous analyses presented in table 6. Results imply that while several variables related to the students’ economic and social contexts are correlated—among them SES, language spoken at home, urban status, and school size—the student’s gender and teacher education are mostly uncorrelated with the other dimensions.

32 Table S2.7 in the supplementary online appendix presents differences in the observable variables related to students, teachers, and parents included in table 2 between teachers with and without university degrees. The only three variables

have stronger effects than teachers with a tertiary, nonuniversity degree. Considering the results in panel A, the effect of the Mimate treatment is 0.29σ for schools in which the treatment was implemented by teachers with university degrees in contrast to an effect of 0.11σ (which is not statistically different from 0) for schools in which the program was implemented by teachers with a tertiary, nonuniversity degree. Strikingly, the interaction between the Mimate treatment and teachers with university degrees has a similar size in both follow-ups (0.18σ in the first follow-up and 0.21σ in the one-year follow-up) and suggests a potential complementarity between Mimate and the teacher's educational level.³³ The inherent classroom difficulties in Peru—large class sizes, limited educational materials, poor infrastructure—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 their less-educated peers. 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. Section 4 presents an additional discussion on this point.

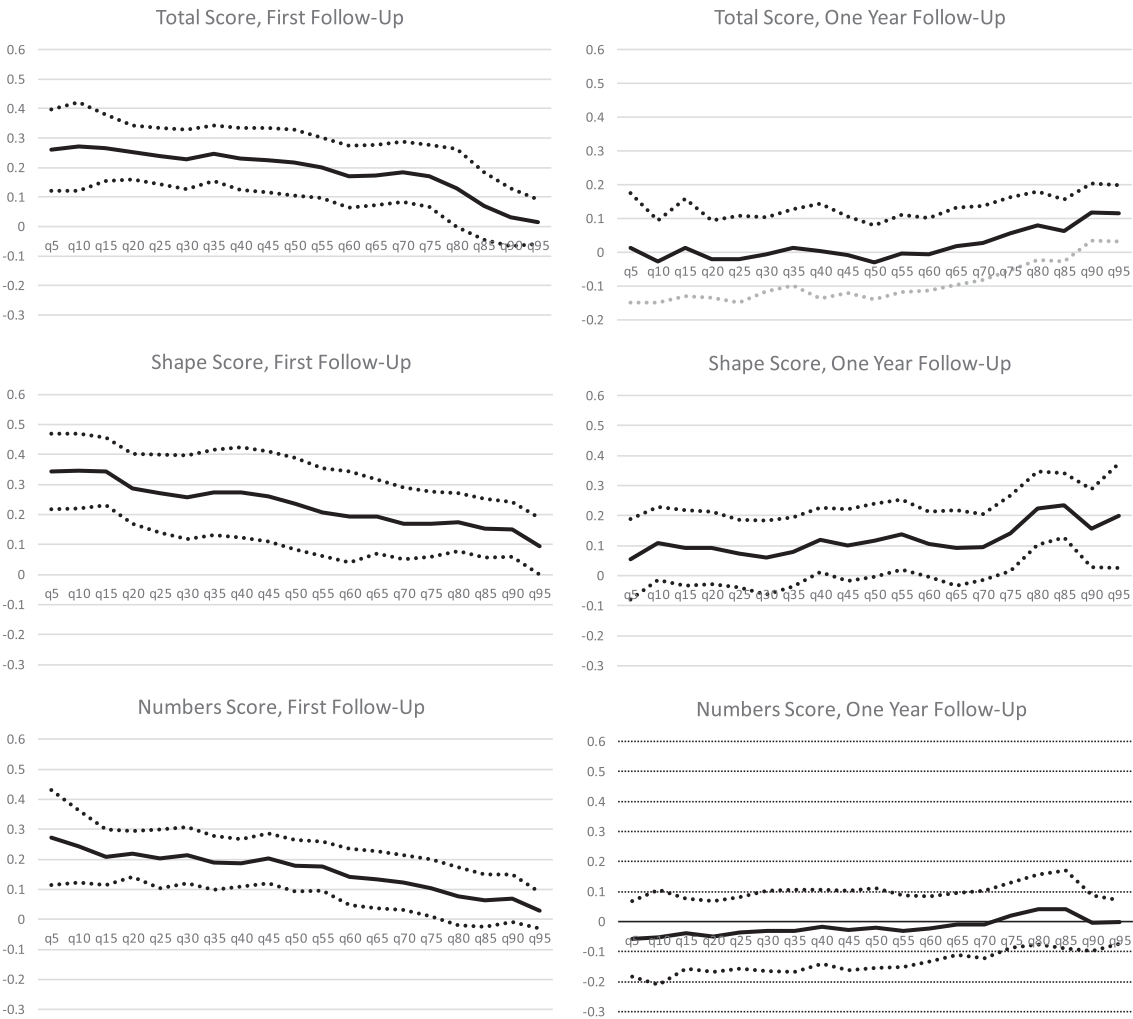
Quantile Regressions

The results of treatment effects using quantile regressions are presented now, including tests of differences across different quantiles of the test score distribution. The analyses estimate treatment effects for the 19 quantiles from quantile 5 to quantile 95.³⁴ Figure 2 reports the results for the complete test, for the shapes items, and for the numeracy items.³⁵ This exercise is important because, in mathematics, teachers often find it particularly difficult to find a balanced lesson plan that stimulates the curiosity of the high-achieving students without confusing the low-achieving ones.³⁶ Mimate's inquiry-based scaffolding model, which adapts to meet the individual needs of students, was designed to address this issue. Results in panel A of fig. 2 imply that for the 5th to 80th quantiles, the estimates for the complete test in the first follow-up are positive and statistically significant (within the 90 percent confidence interval) and range from 0.13σ

that are statistically different among them are age (with teachers with university degrees being younger), language used to teach (teachers with university education are more likely to use both Quechua and Spanish to teach), and the parenting index (with teachers with university degrees having parents with worse parenting practices). This implies that schools with and without university-degree teachers were very similar in most dimensions (21 out of 24). This is most likely a consequence of the fact that the allocations of teachers across public schools in Peru does not make any distinction among teachers with and without university degrees (see “Ley de Carrera Pública Magisterial” #29062 of 2007, which regulated entry into the teaching profession at the time of this study). In turn, table S2.8 in the supplementary online appendix presents balance tables for students with teachers with and without university degrees. As with balance for the complete sample (table 2), the share of girls is not balanced in both samples. In addition, for the sample with teachers with university education, results also show that the share of students who attended previous preschool grades is also larger in treatment students. However, a joint-test of differences across treatment and control groups suggests there is global balance across treatment and control students for this sample (with a p -value of 0.57).

- 33 It is interesting to notice that results show that all the variables for which there are statistical differences among teachers with and without university degrees (years of teaching experience, the parenting index, and language used to teach) and for the unbalanced variables across treatment and control groups for different teachers (gender and whether the student attended previous preschool grades) did not have a significant impact on the effects of the program.
- 34 Figure S2.1 in the supplementary online appendix presents kernel density estimates for math test scores by treatment status in both follow-up surveys. Results suggest that the difference in favor of the treatment group discussed previously is especially relevant for scores located below the average of the control group for the first follow-up survey. These results are suggestive and motivate the quantile regression analyses.
- 35 All regressions control for all the variables included in panel A of table 2. Therefore, estimates are conditional on the student characteristics in the baseline.
- 36 There is a general literature on the effects of several innovations to try to deal with class heterogeneity such as tracking (see Duflo, Dupas, and Kremer 2011 and the references therein) and the use of software that allows students to advance at a different pace (see Barrow, Markman, and Rouse 2009; Muralidharan, Singh, and Ganimian 2018 and the references therein).

Figure 2. Quintile Regression Estimates Panel A: First Follow-Up panel B: One-Year Follow-Up



Source: Data used to construct these figures come from the baseline information and from the first and one-year follow-up tests applied to students. The data were collected by the Peru Office of Innovation for Poverty Action.

Note: Each figure presents treatment effect estimates using quantile regressions (solid line) and their 90 percent confidence intervals (in dashed lines). The estimated quantile regressions control for variables included in panel A of [table 2](#). Estimated standard errors are clustered at the school level using [Parente and Santos-Silva \(2016\)](#).

to 0.27σ . The estimates for the shapes part of the test are statistically significant for all the estimated quantiles, with estimates ranging from 0.10σ to 0.35σ . Estimates for the numeracy part of test are slightly lower in size and only statistically significant for quantiles 5 to 75, with estimates ranging from 0.11σ to 0.27σ .

In terms of gradients across different quantiles, results imply that it is not possible to reject the possibility that the effects are constant across the main part of the distribution of scores except for students located in the top quantiles. For instance, the null hypothesis that the estimates for quantiles 90 and 95 are statistically similar from estimates for quantiles 5 to 70 for the complete math test is rejected. This shows both the adaptability of the program for students at different math outcome levels, but also suggests that the program may be less helpful for students with higher levels of ability, perhaps because they were

already receiving additional attention in the regular classes. To sum up, these results imply an equitable impact of the program in the short term, and suggest that Mimate may even narrow the achievement gap between the highest- and lowest-achieving students.

Panel B of [fig. 2](#) presents the quantile regression results for the one-year follow-up. Treatment effects for students located at different points in the outcome distribution present an interesting pattern. Results show significant effects for the shape tests, which are concentrated among the students located in the top of the distribution (quantiles 75 to 95) and are statistically significant. However, it is not possible to reject the null hypothesis of homogenous effects across the whole distribution of shape scores. In contrast, the effects for numeracy are not different from 0 across students in the complete outcome distribution. In addition, when testing the effects of treatment *across* follow-ups, it is not possible to reject the possibility that the treatment effects are similar in the first and one-year follow-up for quantiles 50 to 95, and that they are statistically different for the other quantiles. This implies that the reduction in treatment effects observed on the shapes items in the one-year follow-up is mostly driven by students in the lower part of the outcome distribution.

These results suggest the existence of different types of complementarities for a program like Mimate. First, there is a static complementarity, where lower-achieving students benefit the most from receiving the program (and, therefore, they suffer the most after the program ends). Second, there is also a dynamic complementarity where, in contrast, the effects of the program over time are higher for high-achieving students.³⁷ This may be a consequence of the intervention only lasting one academic year. More research is needed to untangle the roles that these types of complementarities may be having on the results.

4. Additional Exercises

This section discusses several additional exercises that could shed some light on the mechanisms that may explain the main results. The section also presents a cost-effectiveness analysis of the intervention.

Classroom Observations

This subsection presents results using classroom observations to better understand the mechanisms behind the program's effects. The analyses consider information for 37 treated and 7 control classrooms. This makes it possible to compare in an exploratory way (given the small sample of control classrooms) the practices between treatment and control schools. First, in the area of classroom organization, the classes taught by Mimate teachers were 38 percent more likely to be “prepared and structured with a clear objective” than those of control teachers. For instance, teachers in the control group were observed many times improvising with activities during the lesson. Second, in the area of emotional support, teachers in the treatment group patiently explained mathematics more often than their colleagues in the control group (95 percent vs. 71 percent). Moreover, when students made mistakes, Mimate teachers more consistently encouraged their students—in a friendly manner—to try activities multiple times (in 95 percent of the cases), compared to the control group (71 percent of cases). Third, consistent with the program's inquiry-based approach, teachers in treatment schools often allowed their students to discover the objectives of the activities by themselves (an average of 3.5 in a metric that goes from 0 to 4, with 1 = never and 4 = always). In 100 percent 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 percent) than the control group (63 percent). Thus, the Mimate program seems to have had effects in terms of class management and teacher patience.

37 The article borrows the notation from [Cunha and Heckman \(2007\)](#), where “static complementarity” refers to the effect of the interaction of a program with a feature on current learning, and “dynamic complementarity” refers to the effect of a program with a feature on future learning. The two are conceptually different.

Then the analyses uses the CLASS rubric to compare outcomes between 12 randomly chosen treatment classrooms and 4 control classrooms, also selected at random along the three areas described above (again, the small sample makes these results simply suggestive).³⁸ Results are (1) the average rate of emotional support was higher in treatment schools—5.32 compared with 4.78 in control schools, and this difference is equivalent to about 0.72σ ; (2) the level of classroom organization was higher in the treatment schools—4.47 versus 4.30 in control schools, equivalent to 0.16σ ; and (3) the teaching support was also higher in treatment schools—3.93 versus 3.79 in control schools, equivalent to about 0.15σ .³⁹

In sum, classroom observations show differences in terms of implementation but also give a sense of pedagogical practices in control schools, where instruction seems to be less tailored to students' needs and with lower support for all students.

Teachers' Follow-Up Surveys

Table 7 presents answers from surveys given to 119 teachers in 95 schools as a source of additional information on the program's actual implementation. The analyses present treatment effects for the complete sample and also for teachers with and without a university degree to better understand the factors driving the results (there are 60 teachers with university education; the rest have tertiary, nonuniversity degrees).

Results show several significant differences between the treatment and control groups. Regarding teaching practices and resources,⁴⁰ treatment teachers are more likely to agree that they have had enough time to teach all the required material and to address all aspects of the curriculum. Interestingly, the overall differences in these two dimensions are driven exclusively by large treatment effects for university-educated teachers. This may explain the stronger effects of the Mimate program for these teachers, who could have the human capital needed to apply the program in a way that allows them to optimize their instruction time. Thus, it is not surprising to find treatment effects for teachers with university degrees in their positive feelings toward teaching math in preschool and in their beliefs that students have fun learning math. In contrast, results do not show the same treatment effects for teachers without university degrees.⁴¹ Results also show treatment effects on using games to teach math and on agreeing that they receive enough advice and pedagogical supervision. While both of these effects are slightly larger for teachers with university education, they are not statistically different.

Results also imply improvements in the perceptions that treated teachers have, compared to their control peers, regarding their students' behavior, general abilities to learn and to learn any concept, and math performance. These results are stronger for teachers with university degrees, confirming the previous results. The only dimension in which the effect is larger for treated teachers without university degrees is in student attendance. As previously mentioned, a challenge that all mathematics programs face is related to the potential gender bias. Interestingly, teachers in the Mimate program have more egalitarian beliefs about behaviors in their classroom and their students' abilities. However, as can be seen in panel C of table 7, treatment effects are concentrated among teachers without university degrees (and there are no significant differences in these beliefs among teachers with and without university degrees).

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

39 As a reference, Cruz-Aguayo et al. (2015) report results for the application of the CLASS instrument in K–2 classes in Brazil, Chile, and Ecuador. Their results are similar to what is observed in this study, with relatively lower performance in the area of teaching support and similar results for emotional support and classroom organization.

40 The questions have answers in five categories ranging from strong disagreement to strong agreement, which are coded from 0 to 1 (i.e., 0, 0.25, 0.5, 0.75, 1).

41 It is worth noting that there are few systematic differences among teachers with and without university degrees in terms of the variables included in panel A in table 7. These are the only statistically significant differences: university teachers are (1) less likely to agree with the statement that it is not necessary to do math assessments, (2) more likely to agree that children have fun learning math, (3) more likely to agree with the statement that teaching math requires more than simply knowing the content, and (4) more likely to agree that they should include games in their pedagogical practices.

Table 7. Teacher Survey: Treatment Effects

	Complete sample	Teachers with university degrees	Teachers without university degrees
<i>Panel A: Teaching practices and resources</i>			
I think the way I teach math is the best	0.06	0.08	0.03
You must teach reading and writing before learning math in preschool	0.00	0.04	−0.08
Mathematics is too abstract for children to learn in preschool	0.03	0.06	−0.04
It is enough to teach numbers and shapes in preschool	−0.03	−0.02	−0.03
Free play is the best way for children to learn math	0.05	0.03	−0.01
It is not necessary to do math assessments with children in preschool	0.01	0.03	0.00
Children of preschool age have very basic mathematical skills	0.00	−0.03	0.10
Children have fun learning math	0.05	0.11 *	−0.01
Mathematics tires children	0.04	0.06	−0.01
Mathematics must be taught in preschool	0.04	0.08 *	−0.06
The learning of mathematics is very important for citizenship training	0.05	0.05	0.00
I dedicate enough time to teaching mathematics and cover the objectives	0.11 **	0.15 **	0.03
I feel properly trained to teach math	0.07	0.06	0.06
You need to know more than just the subject to be able to teach mathematics	−0.08	−0.05	−0.14
It is easier for me to teach language than mathematics	−0.01	0.06	−0.15
I have enough time to address all of the curriculum	0.11 *	0.20 **	0.01
I have no difficulties when it comes to structuring math classes	0.05	0.06	0.06
I find it difficult to teach math in preschool	0.01	0.04	−0.04
I would like to learn new ways to teach math	−0.03	0.00	−0.10 *
Teaching math bores me	−0.01	0.01	−0.04
I have the necessary teaching resources	0.07	0.10	0.08
I have adequate advice and pedagogical supervision	0.10 **	0.14 *	0.11
In my classes I include games and activities that involve all students	0.05 *	0.08 ***	0.05
<i>Panel B: Perceptions about students</i>			
My students are well behaved	0.08 **	0.11 *	0.09
My students have the ability to learn	0.10 **	0.15 **	0.10
My students know how to work in a team	0.04	0.02	0.06
My students are very interested in learning	0.04	0.05	0.03
My students have very good attendance	0.03	−0.01	0.12 *
My students have a problem concentrating on their assignments	−0.09	−0.05	−0.16
My students can learn any concept that they are taught	0.10 **	0.15 **	0.03
My students are good at mathematics	0.11 ***	0.15 **	0.03
<i>Panel C: Perception on gender differences</i>			
No gender differences in discipline problems	0.17 **	0.13	0.23
No gender differences in need of help from teacher	0.22 **	0.15	0.24
No gender differences in ability to pay attention in class	0.11	−0.04	0.28 *
No gender differences in ability to follow instructions	0.28 ***	0.16	0.42 ***
No gender differences in mathematics performance	0.26 ***	0.21 *	0.41 ***

Source: Data used in this table come from the baseline information and from the first year follow-up survey applied to teachers. The data was collected by the Peru Office of Innovation for Poverty Action.

Note: Estimates correspond to a test of differences in average for the treatment and control group in each sample. Significance levels: * 10 percent, ** 5 percent, *** 1 percent.

In sum, the information analyzed in this subsection makes it possible to understand the overall impacts and the mechanisms behind them. First, results suggest that the Mimate program helped to improve the management of several classroom dimensions, which allowed teachers to cover more materials and use their time more efficiently. Results also imply improvements in teacher support and supervision, improvements in beliefs about the student’s abilities, and a more gender-neutral approach in treatment schools.

However, the results for teachers with university education (who produced stronger impacts in terms of test scores) suggest that it is likely that the program's impacts could be due to improvements in their classroom management ability than other benefits from the program.

Parents Follow-Up Surveys

This subsection discusses results using surveys for 1,780 parents from 99 schools after the treatment was implemented. The questions were aimed at understanding if the program changed their beliefs and attitudes, how they help with homework, and other parental involvement in their children's education. Results do not imply significant differences between 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 in mathematics (41 percent versus 33 percent in the control group). This suggests that Mimate does not seem to have affected most parental attitudes and practices, and therefore these factors are not driving the observed improvements.

Cost-Effectiveness

This subsection discusses the cost-effectiveness of this intervention. The marginal cost of the intervention is \$37.40 per student (in 2017 U.S. dollars).⁴² Broken down, the educational materials are about \$8 per student, and the remaining cost (about \$29) is the cost of teacher training and support. This does not include the development of the materials used in the program, which had an initial cost of \$200,515 (about \$120 per student). Thus, if considering the marginal cost of expanding the program, the results in this article imply that \$100 increases math performance in about 0.38σ .⁴³ Comparing this to some of the results reported by [Kremer, Brannen, and Glennerster \(2013\)](#) for innovations for primary education in developing countries, results imply that the program is less cost-effective than the most efficient innovations they mention, but is more cost-effective than most of the interventions involving teachers.

5. Discussion and Conclusions

There is no silver bullet that solves all education problems, but tested pedagogical strategies backed by theory are a promising avenue for delivering returns to investments in education ([Duncan and Magnuson 2013](#); [Kremer, Brannen, and Glennerster 2013](#); [Jackson and Makarin 2017](#); [Muralidharan 2017](#)). Moreover, these types of interventions are particularly useful when they allow improvements in educational outcomes without changing the current curriculum and inputs such as teachers and instruction time. This article studies the impacts of a program, Mimate, that was implemented without changing the school curriculum in Peru. The program provides preschoolers with age-appropriate materials that can bolster their mathematical foundations and trains teachers to focus on the needs of the individual student rather than the group and to use interactive instead of rote methods. The importance of mathematics skills at the preschool level has become increasingly apparent, but which specific interventions can improve skills are less known ([Duncan and Magnuson 2013](#); [Dillon et al. 2017](#)). In Peru, as in many developing countries, where underachievement in mathematics persists from preschool all the way to secondary school, a well-designed program could produce a significant effect.

The results in this paper suggest that Mimate has statistically significant impacts after a 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 decrease one year after the program ended but are persistent for items related to shapes. This is consistent with the process

42 If considering the conversion of US\$ to purchasing power parity (PPP), the cost per student is \$61.

43 This is a conservative computation considering an impact of 0.14σ . If the higher estimate of the treatment effect (0.21σ) is used, an \$100 increase in program expenditure would imply a 0.56σ improvement in math test scores.

information presented in this paper, which suggests that the implementation of the program was higher for shapes than numeracy. Results do not imply 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.

The article also finds that while the effects of the program are not significantly different for students with different socioeconomic backgrounds or school types, the effects are much stronger and persistent for teachers with a university education (versus teachers with only a tertiary, nonuniversity degree). This suggests that teachers' human capital complements a program that changes classroom management and personalizes the instruction in an efficient way. This is consistent with information coming from teacher surveys after the program was implemented, which suggest that the program has bigger impacts on the practices and beliefs of these teachers. This information also suggests that a key mechanism to explain the results is the impact of the program on the ability of teachers to manage the classroom more efficiently. Using quantile regressions, results show that the program has stronger effects in the short term for students located in the lower part of the outcome distribution and that the decreased effects in the second follow-up survey is concentrated among those students. This suggests a stronger effect for low-ability students, which does not translate into dynamic advantages. It is not possible to know if this would have been different if the program had been implemented in its totality and not shortened by teacher strikes.

In terms of external validity, it is important to stress that (1) most schools in Peru and other developing countries use the same teaching methods replaced by the Mimate program (e.g., [APOYO 2012](#); [Banerjee et al. 2016](#)); (2) the surveys and class observations collected suggest that teachers, parents, and students liked the program; (3) the process information collected suggests important changes in teacher practices; and (4) the implementation of the program was done using a light-touch approach during regular hours and holding expenditure and staffing constant. These factors suggest that the program can be scaled up as studied. The main limitation is related to teachers' human capital, where the results suggest that the effects are stronger for teachers with university degrees and, therefore, that this type of pedagogical innovation may need teachers with university education to produce large impacts.

There are several implications and questions for future research and experimentation that can be derived from the results presented here. Dramatic changes, like asking experienced teachers to abandon frontal teaching practices and provide inquiry-based and tailored instruction based on each student's learning needs, can take some time to become common practice. The results that more educated—but not more experienced—teachers taught the program more effectively suggest that teaching flexibility, or teaching tools taught in university, are critical to the program's success. In this dimension, teacher-training visits or directed special attention to teachers with lower levels of education could help close this gap. However, further research is 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 as parents would have tangible resources to engage in mathematics with their children.

References

- Almond, D., and J. Currie. 2010. "Human Capital Development before Age Five." *Handbook of Labor Economics* 4B, edited by O. Ashenfelter and D. Card, 1316–476. Amsterdam: Elsevier.
- 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." Unpublished paper, Instituto Apoyo, Lima, Peru.
- Araujo, C., P. Carneiro, Y. Cruz-Aguayo, and N. Schady. 2016. "Teacher Quality and Learning Outcomes in Kindergarten." *Quarterly Journal of Economics* 131 (3): 1415–53.
- Banerjee, A., and E. Duflo. 2010. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. New York: Public Affairs.

- 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, National Bureau of Economic Development, Cambridge, MA.
- Barrow, L., L. Markman, and C. Rouse. 2009. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." *American Economic Journal: Economic Policy* 1 (1): 52–74.
- Berlinski, S., and N. Schady. 2015. *The "Early Years Child Well-Being and the Role of Public Policy."* Washington, DC: Interamerican Development Bank.
- 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, Washington, DC.
- Centro Peruano de Audición, Lenguaje y Aprendizaje (CPAL). 2012. "Informe sobre análisis del instrumento: Evaluación Experimental del Programa MIMATE-Inicial en los Departamentos de Ayacucho y Huancavelica, Perú." Unpublished IADB Consultancy Report, Inter-American Development Bank, Washington, DC.
- Chetty, R., J. Friedman, N. Hilger, E. Saez, D. Schanzenbach, and D. Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4): 1593–660.
- 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." Unpublished paper, Inter-American Development Bank, Washington, DC.
- Cunha, F., and J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.
- de Grave, W. S., D. H. J. M. Dolmans, and C. P. M. van der Vleuten. 1999. "Profiles of Effective Tutors in Problem-Based Learning: Scaffolding Student Learning." *Medical Education* 33 (12): 901–6.
- Dillon, M., H. Kannan, J. Dean, and E. Spelke. 2017. "Cognitive Science in the Field: A Preschool Intervention Durably Enhances Intuitive but not Formal Mathematics." *Science* 357 (6346): 47–55.
- 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.
- Duflo, E., R. Glennerster, and M. Kremer. 2008. "Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*. Vol. 4, edited by T. Schultz and J. Strauss, 3895–962. Amsterdam: North Holland.
- Duncan, G., and K. Magnuson. 2013. "Investing in Preschool Programs." *Journal of Economic Perspectives* 27 (2): 109–32.
- ESCALE. 2019. *Estadística de calidad educativa*. Ministerio de Educación, Peru. <http://escale.minedu.gob.pe/>.
- Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang-Lopez, and S. Grantham-McGregor. 2014. "Labor Market Returns to Early Childhood Stimulation: A 20-year Follow-up to an Experimental Intervention in Jamaica." *Science* 344 (6187): 998–1001.
- Hammil, D. D., N. A. Pearson, and J. K. Voress. 1995. *DTVP: Método de evaluación de la percepción visual de Frostig. Español*. 2nd ed. Mexico City: 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* 3 (2): 97–129.
- He, F., L. Linden, and M. MacLeod. 2009. "A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial." Unpublished paper, Department of Economics, University of Texas, Austin.
- Heckman, J., and S. Mosso. 2014. "The Economics of Human Development and Social Mobility." *Annual Review of Economics, Annual Reviews* 6 (1): 89–733.
- Hogan, K., and M. Pressley. 1997. "Scaffolding Scientific Competencies within Classroom Communities of Inquiry." In *Scaffolding Student Learning: Instructional Approaches & Issues*, edited by K. Hogan and M. Pressley, 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, International Development Bank, Washington, DC.
- Inter-American Development Bank (IADB). 2018. Centro de Información para la Mejora de los Aprendizajes. <https://www.iadb.org/es/cima>.

- Jackson, K., and A. Makarin. 2017. "Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment." *American Economic Journal: Economic Policy* 10 (3): 226–54.
- Kim, J. S., M. A. Burkhauser, D. M. Quinn, J. Guryan, H. C. Kingston, and K. Aleman. 2017. "Effectiveness of Structured Teacher Adaptations to an Evidence-Based Summer Literacy Program." *Reading Research Quarterly* 52 (4): 443–67.
- Kremer, M., C. Brannen, and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.
- Lee, D. 1999. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3): 1071–102.
- Lepper, M. R., M. F. Drake, and T. O'Donnell-Johnson. 1997. "Scaffolding Techniques of Expert Human Tutors." In *Scaffolding Student Learning: Instructional Approaches & Issues*, edited by K. Hogan and M. Pressley, 108–44. 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): 413–28.
- List, J., A. Shaikh, and X. Yu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22 (4): 773–93.
- Lynch, K., and J. S. Kim. 2017. "Effects of a Summer Mathematics Intervention for Low-Income Children." *Educational Evaluation and Policy Analysis* 39 (1): 31–53.
- Muralidharan, K. 2017. "Field Experiments in Education in Developing Countries." In *Handbook of Field Experiments*, Vol. 2, edited by A. Banerjee and E. Duflo, 323–85. Amsterdam: Elsevier, 2017.
- Muralidharan, M., A. Singh, and A. J. Ganimian. 2018. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426–60.
- Näslund-Hadley, E., S. Parker, and J. Hernández-Agramonte. 2014. "Fostering Early Math Comprehension: Experimental Evidence from Paraguay." *Global Education Review* 1 (4): 135–54.
- Parente, P. M. D. C., and J. M. C. Santos Silva. 2016. "Quantile Regression with Clustered Data." *Journal of Econometric Methods* 5: 1–15.
- 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* 59 (1): 187–229.
- Pianta, R. C., K. M. La Paro, and B. K. Hamre. 2008. *Classroom Assessment Scoring System (CLASS) Manual: Pre-K*. Baltimore, MD: Brookes Publishing Co.
- Saye, J. W., and T. Brush. 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): 446–63.
- Stith, A., K. S. Gorman, and N. Choudhury. 2003. "Beyond Primary School: Psychosocial Risk and Gender." *Applied Psychology: An International Review* 52 (4): 614–29.
- Vygotsky, L. 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." *Lancet* 366 (9499): 1804–7.
- Yoshikawa, H., C. Weiland, J. Brooks-Gunn, M. Burchinal, L. Espinosa, W. T. Gormley, J. Ludwig, K. Magnuson, D. Phillips, and M. Zaslow. 2013. "Investing in Our Future: The Evidence Base on Preschool." Society for Research in Child Development, Washington, DC.