JICA Ogata Sadako Research Institute for Peace and Development

JICA Ogata Research Institute Working Paper

Developing Textbooks to Improve Student Math Learning: Empirical Evidence from El Salvador

Takao Maruyama and Takashi Kurosaki

No. 217

January 2021

JICA Ogata Sadako
Research Institute
for Peace and Development

JICA Ogata Sadako Research Institute for Peace and Development
10-5 Ichigaya Honmura-cho
Shinjuku-ku
Tokyo 162-8433 JAPAN
TEL: +81-3-3269-3374
FAX: +81-3-3269-2054

Developing Textbooks to Improve Student Math Learning: Empirical Evidence from El Salvador

Takao Maruyama* and Takashi Kurosaki†

Abstract

School enrollment has rapidly increased since 1990 in developing countries at the primary level but the quality of education has stagnated over the years. In teaching and learning practices, textbooks are an important intermediate that links curriculum, teachers, and students. Since textbooks describe the content and methodology of teaching and learning, they can improve teaching and learning practices, if they are carefully designed. This study evaluates the effectiveness of the package of interventions including the distribution of textbooks that are carefully designed to improve student learning in math through a randomized controlled trial in El Salvador. This experiment tracked same students for two years. The average one-year impact of the package on primary school 2nd grade students' math learning is estimated around 0.48 standard deviation of test scores. The impact was larger on students with higher baseline scores. The average accumulated impact of the first-year interventions one year after is around 0.12 standard deviation. The package of intervention improved math learning of 2nd grade students, and the impact persisted even after schools of the control group also received the package of interventions in the following year.

Keywords: Educational Development; Math textbook development; Math learning; Human Capital and Impact evaluation

---

* Graduate School for Humanities and Social Sciences, Hiroshima University, 1-5-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8529 Japan. (tmaru@hiroshima-u.ac.jp)

† Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan (kurosaki@ier.hit-u.ac.jp).

We are grateful to the Ministry of Education in El Salvador for their understanding and intensive support to conduct this research. We also appreciate Norihiro Nishikata, Kohei Nakayama, Eiichi Kimura, Satsuki Kawasumi, and Yuko Kawanami for their valuable support to conduct the field surveys, and Shin-ichi Ishihara, Hiromichi Morishita, Kazuro Shibuya, Eiji Kozuka, and Chie Esaki for their continuous encouragement for this study. We are grateful to anonymous reviewers for their valuable comments. We also thank Sugashi Nagai, Koei Research & Consulting Inc., and his team members for their dedicated work on data management.

1. Introduction

Approximately 617 million primary and lower secondary school-age children around the world are not reaching the minimum proficiency levels in reading and mathematics (UNESCO 2017). School enrollment has rapidly increased since 1990 in developing countries at the primary level. The adjusted net enrollment rate improved from 80 percent (2000) to 91 percent (2015) (UNDP 2015). In spite of such progress, the quality of education has stagnated over the years. For example, a large volume of second grade students in developing countries including India, Uganda and Nicaragua still cannot solve simple two-digit subtraction (World Bank 2018). In developing countries, more children than in previous years are receiving “schooling” but they are not “learning” well (ibid). Those children who are left behind finish or drop out from primary education without mastering basic competencies. Furthermore, this low level of learning can slow down national economic growth. There is an argument that the Latin American growth puzzle could be resolved by considering achievement of quality education, a direct measure of human capital. The study by Hanushek and Woessman (2012), for instance, revealed that test scores that were larger by one standard deviation were associated with an average annual growth rate in GDP per capita that was about two percentage points higher over 40 years. Thus, improvement of learning is a critical issue for human development and economic development in developing countries.

Numerous interventions for improvement of learning have been evaluated through experimental and quasi-experimental studies (World Bank 2018). One of the examples of interventions successfully scaled up is “Teaching at the Right Level (TaRL)” developed by the Indian NGO “Pratham” (Banerjee et al. 2017). Schools organize simple assessments of basic reading and writing and basic mathematics. Children in grades 3 through 5 are divided into groups by the assessment result, and then they learn through various activities aligned with their proficiency level. The methodology helps those children to rapidly catch up with other children on basic reading and writing, and basic mathematics in a relatively short time (ibid). Those competencies are foundational for children to learn but it is necessary to develop interventions for supporting children to advance their learning after they acquire the basic competencies.

In teaching and learning practices, textbooks are an important intermediate that links curriculum, teachers, and students. Textbooks describe the content and methodology of teaching and learning. While textbooks were regarded as one of the most important inputs to significantly impact student learning up until 2000 (World Bank 2001), recent evidence shows that the magnitude of the impact of the distribution of textbooks on student learning is small (Glewwe 2013). Distribution of textbook only improved learning of students with high baseline scores (Glewwe et al. 2009), and those with high socio-economic status (Kuecken and Valfort 2013). Rapid expansion of school enrollment in developing countries might have brought a mismatch between contents in textbook and student learning level. Snilstveit et al. (2016) argues that it is necessary to address several challenges such as inadequately trained teachers, lack of appropriate materials, curricula, and instructional approaches for improving learning, and mentions that a structured pedagogy program that includes different types of interventions such as teacher training, provision of teaching and learning materials, and mentoring of teachers is effective. While appropriate textbooks can enhance the impact of improvements in pedagogy on student learning, most programs to improve pedagogy do not attempt to develop better textbooks to cover math curricular; thus combining programs to improve pedagogy with the development of better math textbooks is still rare in educational development aid.

To fill in this lack of understanding regarding effective teaching through textbook development, a novel package of interventions has been developed in El Salvador, which we analyze in this paper. In 2015, the Ministry of Education of the country launched the "Project for the Improvement of Mathematics Teaching in Primary and Secondary Education" ("Proyecto de Mejoramiento de Aprendizajes en Matemática en Educación Básica y Educación Media" in Spanish, hereinafter "ESMATE"), with technical cooperation from Japan International Cooperation Agency (JICA). The ESMATE project developed a set of mathematics textbooks, teachers' guidebooks, and student workbooks with an intention to increase on the amount of time students engaged in learning and facilitate teachers to support student learning. In the traditional approach to teaching mathematics in El Salvador, teachers explained how to solve math problems and posed example problems. Students listened to the explanations and took notes, but most of them had difficulty in solving the problems because of their low level of learning. Typically, students merely copied the answer without understanding how it was derived. Teachers did not pay attention to how much students learned, and moved to the next new topic (JICA 2019). The textbooks developed by the ESMATE project, hereinafter referred to as "ESMATE textbooks," intended to change such teaching and learning practices. Subject contents were carefully subdivided considering the student assessment conducted by the project, and the contents were sequenced in the textbook to assure small-step learning by students. The textbooks were designed to facilitate lessons in which students engage in learning mainly through problem solving work.[1] The ESMATE project also designed a package of interventions for schools to improve students' math achievement, hereinafter referred to as "ESMATE programme." The ESMATE programme was composed of (a) distribution of ESMATE textbooks, student workbooks, and teachers' guidebooks, (b) introductory teacher training on the use of the textbook, (c) regular class observations by school principal, (d) review meetings among teachers based on the result of tests, and (e) introductory training of representatives of the parent association.

This study evaluates the effectiveness of the ESMATE programme on 2nd and 7th grade students' math learning outcomes by a randomized controlled trial, and this paper discusses the impact on 2nd grade students.[23] The randomization enables us to precisely identify the causal

---

[1] Textbooks, workbooks, and teachers' guidebooks developed by the project are posted on the following website by the Ministry of Education, Science and Technology in El Salvador.
https://www.mined.gob.sv/materiales-educativos/item/1014902-esmate.html
Sample pages of the textbook and teacher's guidebook are shown in the appendix.

[2] The randomized controlled trial was conducted under the agreement between the Ministry of Education in El Salvador and JICA on June and October 2018. The survey and database construction were done by Koei Research & Consulting Inc., under the contract with JICA, in collaboration with the

impact of the ESMATE programme on students' math learning outcomes, free of selection bias.

This paper makes three main contributions to the literature. First, as the literature is mostly on the impact of distributing existing textbooks, this paper contributes to the extension of the literature with the case for new and improved textbooks structured for changing teaching and learning practices. Providing teaching and learning materials and teacher training does not ensure changes in teaching practices as intended. Thus, the ESMATE programme included other components including lesson observations by the school principal and mutual review meetings among teachers following periodic student assessments. The average one-year impact on 2nd grade student math learning is estimated around 0.5 standard deviation of test scores. This paper further contributes to the extension of related literature in Latin America. While Jamison et al. (1981) investigated the impact of textbooks on learning in Nicaragua, the recent evidence on the impact of textbooks on learning is mainly from Africa. The number of studies on instructional intervention in Latin America is relatively small compared to the other regions (McEwan 2015).

Second, this study evaluates the accumulated impact of the first-year intervention on student math learning in the second year, relative to the impact of receiving the treatment only in the second year. This experiment continued for two years, and tracked the same teachers and students through three rounds of surveys in 2018 (baseline, end-line) and 2019 (follow-up).[4] At the second year, both treatment and control groups received interventions such as grade 3 textbooks. By comparing the learning outcomes of both groups in the second year, we can identify the accumulated impact of the first-year intervention on the following year.[5] Even

---

Ministry of Education in El Salvador. Hitotsubashi University Research Ethics Examination Committee reviewed the research plan. All the data used for this paper is provided by JICA.

[3] The education system of the country is composed of pre-school (age 4 to 6), primary (grade 1 to 6), lower secondary (grade 7 to 9), upper secondary (general: grade 10 to 11, and vocational: grade 10 to 12), and tertiary. The primary and lower secondary education are compulsory. The compulsory education is grouped into 3 cycles; cycle 1 (grade 1 to 3); cycle 2 (grade 4 to 6) and cycle 3 (grade 7 to 9).

[4] The school year in El Salvador starts in mid-January and finishes in mid-November.

[5] We do not refer to the impact as a "two-year impact" since that would refer to the impact on a treatment group that received two years of intervention compared to a control group that received no interventions over two years. In this study, the control group received an intervention in the second year.

though the control group also received a package of interventions in 2019, the impact on student math learning in the treatment group persisted through the end of 2019. The average accumulated impact of the first-year intervention in the following year is estimated to be about 0.12 standard deviation of test scores. The intervention in 2018 (year 1) improved math learning of students in the treatment group; based on the improved learning during year 1, the students in the group could learn new content in the following grade better than those in the control group. Most impact evaluations measure the effectiveness on student learning just after the intervention, but since students continue learning for years, it is important to see the accumulated impact of the first-year intervention on learning in subsequent years.

Third, this paper assesses the heterogeneity of impacts with respect to two aspects that are important in El Salvador. One of them is household assets, whose information we collected through a student interview as part of the baseline survey. In El Salvador, economic inequality is historically high. The Gini index was 54.5 in 1998 and gradually decreased as a result of modest economic growth within the country, reaching 43.4 in 2013 (World Bank 2020). As reported in the following section, the academic achievement of students is moderately correlated with their household economic status in the country. It implies that the economic disparity is reinforced through education over generations. We thus evaluate the heterogeneous impact with respect to baseline scores as well. It turns out that the impact was larger on students with higher baseline scores but the impact does not show a monotonic relation with household economic status. We thus demonstrate that the heterogeneous impact by the baseline score is not brought via the correlation with the student economic status, which is not previously shown in the literature.

The remainder of this paper is organized into four sections: (2) Experimentation Design, (3) Results, (4) Discussion, and (5) Conclusion.

2. Experimentation Design

(1) Contents of package of interventions

The ESMATE project developed a set of mathematics textbooks, teachers' guidebooks, and student workbooks from grade 1 (primary education) through 11 (the last grade of upper secondary education). Subject contents are carefully subdivided considering the student assessment conducted by the project, and the contents are sequenced in the textbook to assure small-step learning by students. The ESMATE textbooks were designed to facilitate lessons in which students engage in learning mainly through problem solving work. This research selected 2nd grade from primary education level, and 7th grade from secondary education level, and evaluated the impact of the ESMATE programme on math learning outcomes using a randomized controlled trial. Considering the difference in the content of interventions and educational levels of those two grades, this paper focuses on primary education and discusses the impact on 2nd grade students.[6]

In 2018, the schools in the treatment group received a package of interventions composed of (a) a set of ESMATE textbooks, student workbooks, and teachers' guidebooks for 2nd grade, (b) introductory training for teachers (b-1), school principals (b-2), and representatives of parent association (b-3), (c-1) math tests aligned with the textbook and (c-2) mutual review meetings among teachers. Introductory training for teachers took place for two days, and include an explanation of the ESMATE textbook and pattern of a lesson within the textbook. In the training, teachers also developed an annual math teaching plan. The plan is a simple one-page year-long calendar that defines which page of textbook will be taught on which day. The school principals were advised to observe math lessons regularly, in total four to five times a year, to give feedback to teachers. Representatives from parent associations participated

[6] In 2018 (first year of evaluation), the ministry distributed ESMATE textbook and teachers' guidebooks nationwide. The evaluation for 7th grade students mainly focused on the impact of the distribution of student workbooks. The result of the experimentation for 7th grade will be discussed in a separate paper.

in a one-day training that focused on the importance of study at home to improve math learning. Mutual review meetings were held in between semesters three times a year, at which teachers brought math test results of their students to review with other colleagues[7]. In El Salvador, lesson observations by school principals and mutual review meetings had been regularly conducted, so the ESMATE project tried to align the content of this existing work to the objective of improving math learning. In 2018, there was no intervention by the ESMATE project to the control group. The survey of this research continued for two years to investigate the accumulated impact of the first-year intervention in the following year. In 2019, the ministry scaled up the package of intervention for all the grades of primary education in public schools across the country including the schools of the control group.[8] The schools in the treatment group also received ESMATE textbooks, teachers' guidebooks, and 3rd grade student workbooks from the ministry.

(2) Sampling

Basic education in El Salvador is divided into three cycles: cycle 1 (grades 1 to 3); cycle 2 (grades 4 to 6) and cycle 3 (grades 7 to 9). Basic education public schools can have preschool, primary, and secondary levels according to the local educational needs. This research targets 2nd grade for primary education and 7th grade for lower secondary education. The sampling frame was composed of basic education public schools offering cycles 1 and 3, according to the targeted grades of this research[9]. There are 14 departments in the country including the capital city, San Salvador. The departments of Cabañas, La Union, San Miguel and San Vicente in the

[7] Mutual review meetings were organized at the prefecture level in this experiment. The meetings were facilitated by the prefectural office of the ministry of education with technical support through the project.
[8] The ESMATE project was finished at the end of June 2019. The ministry integrated the activities in the policy, and allocated necessary budget. Although technical cooperation from JICA was completed, the ministry continued the activities.
[9] Though sampling frame was composed of basic education public schools which had cycle 1 and 3, we eventually found that all the schools also had cycle 2.

central and eastern parts of the country were selected based on their educational statistics such as enrollment and drop-out rates. The educational situation in those four departments is close to or below national averages (Table 1).

Table 1: Basic Educational Statistics in the four departments

	National	Cabañas	La Union	San Miguel	San Vicente
Primary net enrollment rate	86.2%	89.0%	81.2%	85.7%	85.7%
Primary repetition rate (2014)	5.8%	6.7%	5.5%	5.4%	7.7%
Primary drop-out rate (2014)	6.4%	9.8%	8.5%	6.7%	7.7%
Secondary net enrollment rate (2015)	37.9%	25.4%	25.9%	35.5%	38.5%
Secondary repetition rate (2014)	4.9%	3.7%	4.9%	4.2%	4.3%
Secondary drop-out rate (2014)	8.5%	12.4%	11.5%	7.1%	8.0%

Source: Educational statistics of the ministry of education in El Salvador.

In the four departments, there were 1,344 basic education public schools, of which 606 basic education public schools had at least both cycle 1 (grades 1 to 3) and 3 (grades 7 to 9). Around 25 percent of public basic education schools operate with only one or two teachers in the four departments; these were not included in the sampling frame. The country suffers from security problems due to the presence of gangsters inherited from past civil conflicts. Intentional homicides per 100,000 were 61.8 in 2017 (World Bank 2020), the highest in the world. The schools were also affected by gangster activities (USAID 2017). Schools located in areas severely affected by such activities, and any that were physically difficult to access were excluded from the sampling frame. Outside our experimental design, the Millennium Challenge Corporation (MCC) planned to distribute ESMATE textbooks in 2018. The schools receiving intervention from MCC were also excluded from our evaluation framework. As a result, the sampling frame was comprised of 369 basic education public schools. From the sampling frame, 250 basic education public schools were randomly sampled, half of which (125 schools) were randomly assigned to the treatment group while the other half were assigned to the control group (Table 2-1).[10] Stratification variables in the random sampling and treatment assignment of schools were their department and rural/urban designation. If there were several classes of the targeted grades in the sampled school, one class was randomly selected. For security reasons, survey teams conducted field surveys during the morning shift at all schools that offered one, or in the afternoon if necessary. In the baseline survey, 7 schools in the treatment group and 4 schools in the control group were excluded for security reasons (Koei Research & Consulting Inc. 2018).[11] In addition to these eleven excluded schools, there were no students enrolled in grade 2 at one school in the treatment group.

Based on the educational census survey data from the ministry of education in El Salvador, we tried to compare the sampling frame with the original population of schools in the four departments (Table 2-2). Because the data on some schools was not available in the educational census survey data, the number of schools in Column (A) to (C) in Table 2-2 does not exactly match that in Table 2-1. Since this research sampled schools that offer cycles 1 and 3, the size of schools in the sampling frame (Column (C) in Table 2-2) is larger than the population of schools in the four departments (Column (A) in Table 2-2). The percentage of schools in urban area is larger in the sampling frame than the original population of schools. The percentage of schools in the sampling frame with facilities such as libraries is also slightly larger than the population of schools in the four departments. In Table 2-2, we compared characteristics of the original sample of 250 schools and the remaining 238 schools after attrition of 12 schools.

[10] We calculated the sample size with the following conditions: minimum detectable effect size: 0.2 standard deviation of test scores; cluster size: 20 students on average; significance level: 0.05; power: 0.8; and intra-cluster correlation coefficient: 0.25. Considering the risk of attrition of schools because of security issues, 12 schools were added respectively to the treatment and the control groups. The actual value of intra-cluster correlation coefficient of the end-line scores is around 0.30.

[11] At the end-line survey, 3 schools in the control group were additionally excluded because of security reasons (Koei Research & Consulting Inc. 2019).

The characteristics of remaining 238 schools are equivalent to the original sample of 250 schools[12].

Characteristics of students, teachers, and schools in the treatment and control groups are presented in Tables 3-1 to 3-3. The tables show that the treatment and the control groups are well balanced, indicating successful randomization.

Table 2-1: Sampling frame of schools in the four departments

	Cabañas	La Union	San Miguel	San Vicente	Total
(1) No. of public schools (primary and/or lower secondary)	265	375	468	236	1,344
(2) Schools with cycle 1 through cycle 3 in (1)	104	144	247	111	606
(3) Schools without difficulty in access or security in (2)	64	68	164	105	401
(4) Schools not targeted by the MCC program (Sampling frame)	64	49	151	105	369
(5) Sampled schools (Total)	43	33	102	72	250
(6) Sampled schools (Treatment)	22	16	51	36	125
(7) Sampled schools (Control)	21	17	51	36	125

Note: Data sources are baseline survey of this research, and the educational statistics of the ministry of education in El Salvador.

[12] According to the educational census data collected by the ministry of education, there were no statistically significant differences between public schools with cycle 1 & 3 in the four departments (column (B) in Table 2-2) and public schools in the sampling frame (column (C) in Table 2-2).

Table 2-2: Comparison of characteristics of schools

Content	Public schools in four departments (A)	Public schools that have cycle 1 and 3 (B)	Sampling frame (C)	Sample (D)	Surveyed schools (E)	P-value (D)=(E)
Percentage of N. of schools in urban area (2018)	18.6	25.0	29.4	29.2	30.7	0.72
Average N. of students (grade 2) (both shifts) (2018)	14.2	24.0	25.1	26.0	26.4	0.87
Average N. of total students (grade 1 to 9) (both shifts) (2018)	116.3	216.4	227.7	233.7	238.2	0.82
Percentage of Male students in grade 2 (2018)	51.9	52.0	51.7	51.4	51.4	0.99
Percentage of grade 2 students in morning shift (2018)	82.7	80.0	79.6	80.5	79.8	0.81
Percentage of grade 2 students repeated (2018)	3.5	3.2	3.0	3.2	3.2	0.99
Average age of grade 2 students (2018)	8.9	9.0	9.2	8.6	8.6	0.93
School infrastructure: electricity (2016)	97.0	99.0	98.9	98.8	98.7	0.95
School infrastructure: Water (2016)	74.2	82.5	82.9	83.6	84.5	0.8
School infrastructure: Computer (2016)	58.9	77.2	80.7	80.0	80.2	0.94
School infrastructure: Internet (2016)	23.7	37.4	40.0	41.2	42.0	0.86
School infrastructure: Library (2016)	15.3	23.0	24.2	24.0	24.8	0.84
School infrastructure: Laboratory (2016)	6.1	8.6	9.2	9.6	9.7	0.98
School infrastructure: Kitchen (2016)	76.3	79.0	78.3	80.0	79.4	0.87
N. of schools (of which census survey data is available)	1342	605	368	250	238	
N. of schools (which have grade 2 in 2018, and whose census survey data is available)	1226	601	364	249	238	

Note: (1) Data source is educational census survey data in El Salvador. Because the data of some schools are not available in the census survey data, the numbers of schools in Column (A) to (C) in this table do not exactly match with those in Table 2-1. (2) 10% significance: *, 5% significance: **, 1% significance: ***. (3) Values on school facilities are binary (Yes:1, No:0). (4) The p-values on number of students, percentage of students in morning shift and percentage of students who repeated 2nd grade are the results of Wilcoxon rank sum test with stratified data (department dummy, urban/rural dummy). (5) The p-values on binary values show the results of chi-squares test with stratified data (department dummy, urban/rural dummy).

Table 3-1: Characteristics of students (baseline)

Content	Treatment	Control	Mean Diff.	Adjusted Mean Diff. (a)	Standard Error of (a)	P-Value of (a)
Morning Shift (%)	94.94	91.22	3.72	4.71	3.10	0.12
Age	7.83	7.71	0.07	0.05	0.04	0.22
_sd	0.83	0.79				
Sex (Male) (%)	49.61	51.62	-2.01	-2.35	2.05	0.25
N. elder brother/sister	1.62	1.57	0.04	0.05	0.07	0.51
_sd	1.83	1.79				
N. younger brother/sister	0.83	0.82	0.01	0.01	0.03	0.77
_sd	0.94	0.89				
Test score						
Raw test score (Total points: 20)	5.05	4.75	0.30	0.39	0.27	0.14
_sd	3.41	3.65				
Raw test score (Total points: 18, which excludes Q2&Q4)	3.72	3.53	0.20	0.27	0.23	0.24
_sd	3.00	3.18				
Asset of study						
Math textbook 2017 (%)	30.44	35.97	-5.53	-5.89	3.40	0.08 *
Math notebook 2017 (%)	87.36	87.81	-0.45	-1.65	2.64	0.53
Notebook only for Math 2017 (%)	80.60	80.99	-0.39	-1.73	2.86	0.54
Own Study Desk at Home (%)	32.18	33.21	-1.03	-0.72	2.33	0.75
Asset of student household						
Smartphone (%)	74.87	74.81	0.06	-0.17	2.48	0.94
Computer (%)	24.77	22.48	2.29	1.84	1.81	0.30
Refrigerator (%)	82.77	80.88	1.89	1.93	1.76	0.27
Car (%)	31.79	30.61	1.18	0.78	2.02	0.69
TV (%)	90.92	91.87	-0.96	-1.09	1.24	0.38
Tap water (%)	79.41	80.28	-0.87	-0.84	2.27	0.71
Electricity (%)	95.15	94.42	0.73	0.67	1.29	0.60
Flush Toilet (%)	56.09	52.38	3.71	3.31	2.83	0.24
Using wood for cooking (%)	57.33	58.94	-1.61	-1.26	2.96	0.67
Using gas for cooking (%)	89.73	89.76	-0.03	-0.13	1.28	0.91
Using electricity for cooking (%)	6.09	4.50	1.59	0.89	1.60	0.57
N. of schools	117	121				
N. of students	1939	1846				

Note: (1) Data source is baseline survey of this research. (2) 10% significance: *, 5% significance: **, 1% significance: ***. P-value of adjusted mean difference in this table is obtained by regressing the value of each characteristics on treatment assignment dummy with controlling stratification variables (department and urban/rural dummies, and the interactions). Robust standard errors are clustered at school level. (3) Binary values are Morning Shift (Morning shift: 1), Sex (Male:1), Textbook / Notebook (Yes:1, No:0), and Asset of study and student household (Yes:1, No:0).

Table 3-2: Characteristics of teachers (baseline)

Content	Treatment	Control	Mean Diff.	P-Value
Sex	0.22	0.25	-0.03	0.64
Age	45.93	47.10	-1.17	0.39
_sd	9.35	7.82		
Total teaching period (years)	22.26	22.59	0.00	0.86
_sd	8.40	8.22		
Academic Degree				
High school	0.07	0.04	0.03	0.35
Professorate	0.71	0.75	-0.04	0.45
Bachelor	0.21	0.19	0.02	0.77
Master	0.00	0.00	0.00	0.79
Doctor	0.00	0.00	0.00	0.79
Teacher qualification (1)				
Pedagogical Bachelor	0.14	0.07	0.07	0.07 *
Professor	0.74	0.82	-0.07	0.16
License in Education	0.15	0.10	0.05	0.27
Master's in Education	0.00	0.00	0.00	0.79
Doctorate in Education	0.00	0.00	0.00	0.79
Pedagogical Training Course	0.02	0.05	-0.03	0.16
Teacher qualification (2)				
Basic Education Teacher (Cycle I and II)	0.47	0.51	-0.04	0.51
Mathematics Specialty Teacher (Cycle III and High School)	0.08	0.07	0.00	0.94
Teacher specialized in other than math (Cycle III and High School)	0.20	0.23	-0.03	0.51
Class facility				
Board	0.99	1.00	-0.01	0.3
Teacher's desk	1.00	0.97	0.03	0.04 * *
Teacher's chair	0.91	0.88	0.03	0.46
File cabinet / shelves	0.71	0.69	0.02	0.69
Working condition				
Teaching grade (Both 2nd and 7th grades)	0.10	0.09	0.01	0.76
Teaching other subject	0.89	0.90	-0.01	0.76
Only morning shift	0.86	0.87	0.00	0.91
Only afternoon shift	0.06	0.08	-0.02	0.49
Both shifts	0.08	0.05	0.03	0.38
N of schools	117	121		

Note: (1) Data source is baseline survey of this research. (2) 10% significance: *, 5% significance: **, 1% significance: ***. (3) Values on class facilities are binary (Yes:1, No:0). (4) The p-values on Age is the result of Wilcoxon rank sum test with stratified data (department dummy, ur- ban/rural dummy). (5) The p-values on binary values show the results of chi-squares test with stratified data (department dummy, urban/rural dummy).

Table 3-3: Characteristics of schools (baseline)

Content	Treatment	Control	Mean Diff.	P-Value
Number of students				
N of Student (2nd grade) Morning Shift	24.60	21.98	2.62	0.48
sd	18.06	14.03		
N of Student (2nd grade) Afternoon Shift	24.38	19.64	4.74	0.3
sd	14.12	8.27		
N of Student (Total)	259.48	235.27	24.21	0.73
sd	231.99	176.95		
Repetition and Drop out rate				
Repetition rate (morning shift of 2nd grade in 2017)	0.03	0.03	0.01	0.2
Repetition rate (afternoon shift of 2nd grade in 2017)	0.04	0.03	0.00	0.98
Dropout rate (morning shift of 2nd grade in 2017)	0.05	0.06	-0.01	0.85
Dropout rate (Afternoon shift of 2nd grade in 2017)	0.08	0.04	0.03	0.26
N. of teachers				
N. of vice school principle	1.10	1.13	-0.03	0.8
_sd	0.52	0.60		
N. of teachers	10.66	10.05	0.61	0.84
_sd	9.22	6.82		
School facility				
Electricity	1.00	1.00	0.00	0.79
Drinking Water	0.80	0.86	-0.06	0.24
Computer	0.95	0.93	0.02	0.46
Internet	0.44	0.48	-0.04	0.5
Internet_use_for_pupils_students	0.67	0.71	-0.04	0.65
Library	0.26	0.25	0.01	0.88
Laboratory	0.12	0.07	0.05	0.15
Kitchen	0.78	0.81	-0.03	0.54
Canteen	0.81	0.72	0.09	0.09 *
Student support				
School lunch	0.97	0.97	0.00	0.96
Supplementary class (math)	0.23	0.16	0.07	0.15
Donor support within 5 years (except ESMATE)	0.91	0.93	-0.02	0.57
N of schools	117	121		

Note: (1) Data source is baseline survey of this research. (2) 10% significance: *, 5% significance: **, 1% significance: ***. (3) Values on school facilities are binary (Yes:1, No:0). (4) The p-values on number of students, repetition rate and dropout rate are the results of Wilcoxon rank sum test with stratified data (department dummy, urban/rural dummy). (5) The p-values on binary values show the results of chi-squares test with stratified data (department dummy, urban/rural dummy).

(3) Assessment of math learning level

To assess the math learning of students, we conducted written tests in all three rounds of surveys (baseline, end-line, and follow-up). The baseline survey was conducted from January to March 2018, the end-line survey from September to October 2018, and the follow-up survey from September to October 2019. The school year in El Salvador starts in mid-January and finishes in mid-November. To account for the progress following the curriculum, test items differ across these three tests. The tests were designed to measure student learning of math content defined in the curriculum. The test administered during the baseline survey assessed math content learned in the 1st grade. The test given at the end-line survey assessed math content learned in the 2nd grade. Each of these tests consisted of 20 questions, including problems posed in texts. The test given with the follow-up survey assessed math content learned in the 3rd grade. This test consisted of 25 questions, 5 of which were the same as questions given at the end-line survey.[13] Each test assessed different cognitive skills (knowing, applying, and reasoning) and cognitive domains (number and operation, quantity and measurement and geometry). The number of items per unit of math content was defined according to the volume of lesson hours necessary for the unit. The compositions of test items were presented in Tables A-1 to A-3 in appendix 2.[14] The duration of the math tests was 45 minutes. The tests were administered by survey teams without the presence of teachers in the class room.

(4) Student attrition

The majority of teachers and students included in the sample were tracked through three rounds of surveys in 2018 (baseline and end-line) and 2019 (follow-up). Some 2nd grade students moved to different schools and some dropped out of school. In these cases, we were not able to

[13] All the test items are different between the baseline and end-line tests.

[14] In the baseline test, 4 out of 20 questions were multiple choice and the others were open-ended responses. In the end-line test, 2 out of 20 questions were multiple choice and the others were open-ended. In the follow-up test, 1 out of 25 questions was multiple choice and the others were open-ended. For more detail, please refer Tables A-1 to A-3.

track them, resulting in attrition in the dataset. As shown in Tables 4-1, the attrition occurred at similar rates for treatment and control groups.

Table 4-1: Sample attrition

	N. of students	Change from baseline	% of change from the original sample
Treatment Group			
Baseline	1,939		
End-line	1,579	-360	18.6%
Follow-up	1,487	-452	23.3%
Control Group			
Baseline	1,846		
End-line	1,453	-393	21.3%
Follow-up	1,386	-460	24.9%

Note: Data sources are baseline, end-line and follow-up surveys of this research.

We checked whether differential attrition occurred between the treatment and the control groups by regressing the student attrition dummy on the treatment assignment dummy, student characteristics, and stratification variables (department, urban/rural dummies and the interactions). The results, shown in Table 4-2, indicate that the attrition was not differential between the two groups.[15] The results also show that attrition tended to occur for students with lower academic achievement and higher age.

[15] We checked the balance of student characteristics in the samples who remained at the end-line and follow-up survey respectively, and confirmed the balance between the two groups. The results of the comparison are shown in Table B-1 and B-2 in the appendix.

Table 4-2: Sample attrition analysis

	OLS End-li (I)	Logit End-li (II)	OLS Follow-u (III)	Logit Follow- (IV)
(Intercept)	-0.049		-0.034	
	(0.109)		(0.095)	
Treatment	-0.024	-0.023	-0.016	-0.015
	(0.025)	(0.024)	(0.021)	(0.021)
La Union	-0.012	-0.013	0.042	0.046
	(0.038)	(0.045)	(0.043)	(0.048)
San Miguel	0.095^{**}	0.100^{**}	0.063	0.065
	(0.043)	(0.051)	(0.038)	(0.041)
San Vicente	0.055	0.063	0.027	0.029
	(0.039)	(0.049)	(0.033)	(0.037)
Urban	0.021	0.013	0.056	0.058
	(0.041)	(0.055)	(0.040)	(0.044)
La Union×Urban	0.033	0.046	-0.050	-0.049
	(0.057)	(0.079)	(0.064)	(0.056)
San Miguel×Urban	-0.012	-0.005	-0.033	-0.035
	(0.059)	(0.065)	(0.054)	(0.051)
San Vicente×Urban	0.008	0.016	-0.038	-0.038
	(0.061)	(0.074)	(0.052)	(0.051)
Z score baseline (2)	-0.039^{***}	-0.042^{***}	-0.040^{***}	-0.043^{***}
	(0.008)	(0.009)	(0.008)	(0.009)
Age	0.042^{***}	0.036^{***}	0.036^{***}	0.033^{***}
	(0.010)	(0.008)	(0.010)	(0.008)
Sex	0.018	0.018	0.006	0.006
	(0.013)	(0.013)	(0.013)	(0.013)
No. of elder brother/sister	0.005	0.005	0.005	0.005
	(0.004)	(0.004)	(0.004)	(0.004)
N. of younger brother/sister	-0.003	-0.004	0.004	0.004
	(0.006)	(0.006)	(0.007)	(0.007)
N. of asset types	-0.010^{*}	-0.009^{*}	-0.005	-0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Shift (Morning=1)	-0.112^{*}	-0.116^{*}	-0.042	-0.042
	(0.056)	(0.062)	(0.033)	(0.034)
R^2	0.036		0.020	
Adj. R^2	0.033		0.016	
Num. obs.	3,783	3,783	3,783	3,783
F statistic	6.466		4.531	
N Clusters	238		238	
Log Likelihood		-1818.62		-2043.57
Deviance		3637.25		4087.15

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data sources are baseline, end-line and follow-up surveys of this research.
(2) Dependent variable is a dummy that takes 1 for students who were absent respectively at the end-line or follow-up survey. Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(3) Coefficients of the logit regression show marginal effects.
(4) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

Looking at the school-level drop-out rates during the period of impact evaluation, we also found similar rates between schools in the treatment group and schools in the control group. In 2018, the drop-out rate of 2nd grade students in treatment schools was 6.77 percent, and that of the control group was 7.06 percent. In 2019, the drop-out rate of the 3rd grade students in treatment schools was 8.01 percent, and that of the control group was 8.30 percent.

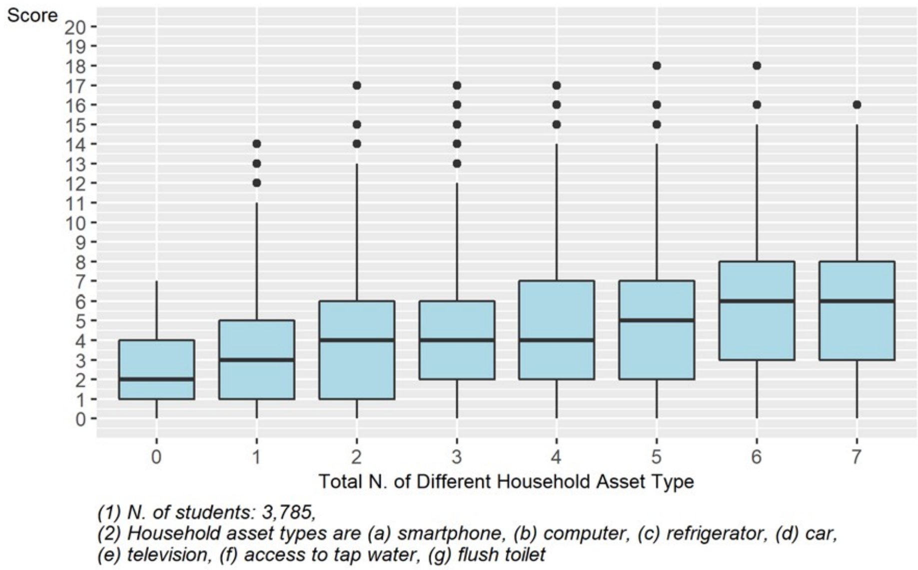
Those 2nd grade students who repeated the same grade in the following school year were kept in our sample and given them the same math test for the 3rd grade students in the follow-up survey. In the follow-up survey data, 1.68 percent of students in the treatment group and 1.3 percent of students in the control group repeated the 2nd grade.

(5) Baseline test scores

At the beginning of the 2018 school year, we conducted the baseline survey of this study, in which 3,785 2nd grade students were tested on math content learned in the 1st grade. The test result shows that most of 2nd grade students did not master the basic mathematic understanding and skills supposed to be learned in the previous grade. For example, while around 80 percent of students correctly responded to single-digit addition problem “1+3”, the percentage of students who could correctly answered single-digit subtraction problem “5-3” was around 25 percent. For the item which asks students to count the number of 17 circles written in the test sheet, the correct response rate was around 50 percent. Only 25 percent of students could identify a figure of rectangle among three figures (circle, square and rectangle).

The baseline survey collected individual student data on different types of household assets: (a) smartphone; (b) computer; (c) refrigerator; (d) car; (e) television; (f) access to tap water; and (g) flush toilet. Figure 1 is a boxplot of 2nd grade student test scores by the total number of different types of student household assets. The total number indicates student economic status. The graph shows that students with higher economic status tend to obtain higher math scores.

Figure 1: Boxplot of the baseline test score of 2nd grade student by total number of household asset types



Note: Data source is baseline survey of this research.

Because of logistical reasons, we were not able to conduct the baseline survey before any component of the intervention package started. The baseline survey was started in mid-January 2018, just after the distribution of textbooks to students at the beginning of a school year (the survey started on January 20 and finished on March 1, 2018). The survey in treatment and control schools was conducted in parallel. The balance of baseline scores is checked using the regression equation

$$Y_{ijk0}=\alpha_0+\delta_0 Treatment_k + D_k\beta_{0D}+\varepsilon_{ijk0} \quad (1)$$

where Y_{ijk0} represents the math test baseline score for student *i* in grade *j* in school *k*. Test scores are standardized by mean and standard deviation of the scores among students belonging to the control group. *Treatment* is an assignment to the treatment group. D_k is a vector of stratification variables in the random sampling and treatment assignment, i.e., department dummies and the rural / urban dummy of school *k* and the interactions. Robust standard errors are clustered at the school level. The result is presented in Table 5. The estimated value of δ_0 is positive but not statistically significant, indicating the treatment and control groups are well balanced.

Table 5: Comparison of baseline score

	Z score baseline	Z score baseline excluding Q2&Q4
(Intercept)	-0.148^{*}	-0.147^{**}
	(0.075)	(0.074)
Treatment	0.106	0.084
	(0.073)	(0.072)
La Union	-0.131	-0.113
	(0.149)	(0.153)
San Miguel	-0.141	-0.128
	(0.099)	(0.095)
San Vicente	0.152	0.138
	(0.111)	(0.111)
Urban	0.667^{***}	0.664^{***}
	(0.184)	(0.199)
La Union×Urban	-0.456^{*}	-0.483^{*}
	(0.253)	(0.264)
San Miguel×Urban	-0.397^{*}	-0.400^{*}
	(0.218)	(0.228)
San Vicente×Urban	-0.295	-0.290
	(0.242)	(0.252)
R^2	0.070	0.065
Adj. R^2	0.068	0.063
Num. obs.	3,785	3,785
N Clusters	238	238

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data source is baseline survey of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.

It would be plausible to think that the intervention might have already affected student math learning measured in the baseline because it was conducted just after the distribution of textbooks. The kernel density curves of the baseline scores (Figure 2-1&2-2) show that the students with the lowest quartile baseline scores in the treatment group obtained slightly better scores than those in the control group. There were 2 items (question No. 2 and No. 4) that students with the lowest quartile baseline scores in the treatment group correctly answered more successfully. Thus, this study also uses an alternative definition of the baseline score that excludes these 2 items.

Figure 2-1: Kernel density curves of Z score (20 test items) at the baseline

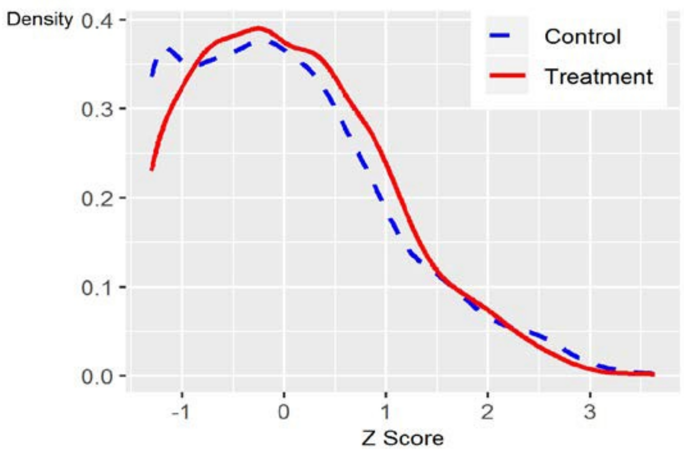
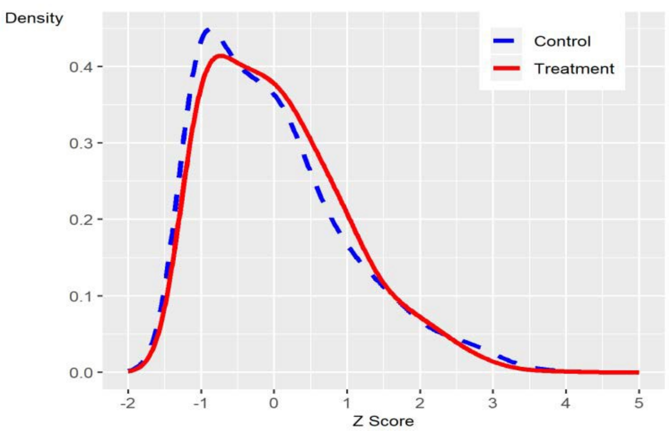


Figure 2-2: Kernel density curves of Z score (18 test items) at the baseline



Note (Figure 2-1, 2-2): Data source is baseline survey of this research.

(6) Estimation strategy[16]

The impacts of the ESMATE programme on student math learning in year 1 and year 2 will be estimated by the following regression equation (2)

$$Y_{ijkt}=\alpha_t+\gamma_t Y_{ijk0}+\delta_t Treatment_k+ C_{ijk}\beta_{tc}+ P_{mjk0}\beta_{tp}+ S_k\beta_{ts}+ D_k\beta_{tD}+\varepsilon_{ijkt} \quad (2)$$

where Y_{ijkt} represents the math test score for student *i* in grade *j* in school *k* at the round *t* of survey (t=0: baseline, t=1: end-line (year 1), and t=2: follow-up (year 2)). Test scores are standardized by mean and standard deviation of test scores of students belonging to the control group at each round of the survey. Sub-totals of test scores by cognitive skills and domains are also used. C_{ijk} is a vector of characteristics of student *i* in grade *j* at school *k* such as age, gender, shift at school (morning or afternoon),the number of brothers and sisters, and characteristics of family of student *i* such as the number of household asset types at the baseline. P_{mjk0} is a vector of characteristics of teacher *m* of grade *j* at school *k*, who teaches mathematics to student *i* in year 1, such as age, gender, and educational qualification at the baseline. S_k is a vector of characteristics of school *k* such as the number of students, school infrastructure, school meal and remedial activities, and characteristics of the school principal. D_k is a vector of department dummies and the rural / urban dummy of school *k* and the interaction terms. Robust standard errors are clustered at the school level.

We do not estimate equation (2) using three-period panel data. Instead, we estimate it using cross-section data composed of the control variables of t=0 and the dependent variable of t=1 for identifying the treatment effect in year 1, or cross-section data composed of the control variables of t=0 and the dependent variable of t=2 for identifying the accumulated treatment effect of year 1 in the following year. Therefore, δ_t in equation (2) is a scalar, not a vector of

[16] The pre-analysis plan of this research was registered at the following website on October 2018: https://www.socialscienceregistry.org/trials/3169. This study follows the estimation strategy defined in the plan. Additionally, this study conducted exploratory analysis in terms of the average impact of the ESMATE programme brought by the difference of textbook in the section 3. (1).

parameters. Both cross-section data are constructed from the balanced panel data either of t=0 and t=1 or t=0 and t=2. Even in the case that the one-year impact on student math learning is positive, the difference of student test scores between the two groups might have vanished at the end of the following year, since students in the control group also received the ESMATE textbooks in year 2.

This study also investigates the heterogeneity of impact on student math learning outcomes by baseline scores and the total number of different types of student household assets using the following regression formula. Letting X_{ijk0} stand for either the baseline score or the total number of different asset types, the estimated model becomes

$$Y_{ijkt}=\alpha_t+\gamma_t Y_{ijk0}+(\delta_{tA}+ \delta_{tB} X_{ijk0})\ Treatment_k+ C_{ijk}\beta_{tc}+ P_{mjk0}\beta_{tp}+ S_k\beta_{ts}+ D_k\beta_{tD}+\varepsilon_{ijkt} \quad (3)$$

If δ_{tB} in equation (3) is positive, the impact of the treatment is larger on students with a higher baseline score or higher economic status. As the baseline score and the economic status are correlated, equation (4) is also estimated, including both of the interaction terms. Equation (5) adds the triple interaction term of treatment assignment, baseline score, and economic status to the equation:

$$Y_{ijkt}=\alpha_t+\gamma_t Y_{ijk0}+(\delta_{tA} + \delta_{tB} Y_{ijk0} + \delta_{tC} Asset_{ijk0})Treatment_k+$$
$$C_{ijk}\beta_{tc}+ P_{mjk0}\beta_{tp}+ S_k\beta_{ts}+ D_k\beta_{tD}+\varepsilon_{ijkt} \quad (4)$$

$$Y_{ijkt}=\alpha_t+\gamma_t Y_{ijk0}+[\delta_{tA} + \delta_{tB} Y_{ijk0} + \delta_{tC} Asset_{ijk0}+ \delta_{tD}(Y_{ijk0} \times Asset_{ijk0})]Treatment_k+$$
$$C_{ijk}\beta_{tc}+ P_{mjk0}\beta_{tp}+ S_k\beta_{ts}+ D_k\beta_{tD}+\varepsilon_{ijkt} \quad (5)$$

where $Asset_{ijk0}$ is the total number of asset types of student *i* in grade *j* at school *k* at the baseline. If δ_{tB} is positive and δ_{tC} is close to zero and not statistically significant, it indicates that the heterogeneous impact on the higher baseline score is not brought about by the student's economic status. The coefficient δ_{tD} represents the multiplier impact by baseline score level and economic status.

3. Results

(1) The one-year impact of a package of interventions for 2nd grade students[17]

Kernel density curves of Z scores at the end-line survey (Figure 3) show that the 2nd grade students in the treatment group as a whole improved math learning. The regression results from equation (2) applied to the t=0 and t=1 panel data are presented in Table 6-1. The average impact of treatment is estimated at around 0.48 standard deviation (Model 2-(5), Model 2-(6)), and statistically significant at the 1 percent level.[18] Since the attrition rate in the control group was slightly higher than the treatment group at the end-line survey, we estimated "Lee bounds" by trimming around 2 to 3 percentage points of students either with the highest score or the lowest score at the end-line survey in the treatment group. The estimated bound of the impact is from 0.41 standard deviation (standard error: 0.067) to 0.53 standard deviation (standard error: 0.070).

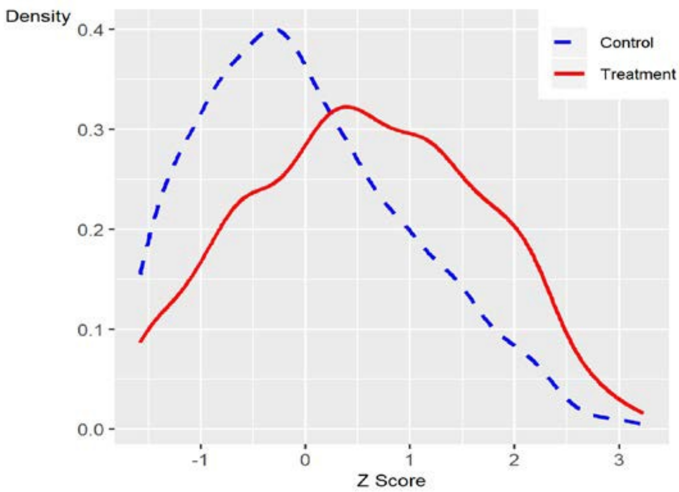
We conducted cost-effective analysis, following the methodology presented by J-PAL (Bhula, R. et al. 2020; Dhaliwal et al. 2014). The cost-effectiveness is measured as the ratio of the aggregated impact of the project (the average impact on student learning per student multiplied by the number of students impacted) to the aggregated cost of implementing the project. The cost-effectiveness is presented as the total standard deviations gained across the sample per 100 USD spent. While travel allowances were not provided by the ministry for teachers at the introductory training, we included the cost for comparison of cost effectiveness with other cases. The cost-effectiveness of the ESMATE programme, the total standard

[17] Analysis in terms of the average impact of the ESMATE programme brought by the difference of textbooks in this section is exploratory.
[18] Considering student data attrition, we also conducted regression analysis of the impact with a weighted sample. The sample was weighted with inverse probability weighting. The alternative result was almost the same as the values reported in Table 7-1.

deviations gained across the sample per 100 USD spent, is estimated to be 3.98. The level of cost-effectiveness of the ESMATE programme is comparable to or higher than the other similar programs cited in Kremer et al. (2013).[19]

Regardless of the cognitive skills measured in assessment tests, the impact was positive and statistically significant, but the magnitude of impacts is largest in the skill of knowing, followed by applying, and then reasoning (Table 6-2). When we distinguish the group of questions by their cognitive domains, the impacts on number and operation are positive and statistically significant, while the impacts on quantity and measurement are insignificant (Table 6-3). The domains of quantity and measurement were assessed using time and amount of water. The result indicates that it would be necessary to check whether textbooks were sufficient for students to learn the topics well for further improvement.[20]

Figure 3: Kernel density curves of Z score at the end-line survey



Note: Data source is end-line survey of this research.

[19] Among 27 programs cited in Kremer et al. (2013), there are five programs whose cost-effectiveness is higher than the ESMATE programme. Though the level of cost-effectiveness varies among the programs, the cost-effectiveness of the textbook program in Kenya for the top 20 percent of students is estimated at 3.56.

[20] In addition to the textbook, it might be necessary to provide additional learning materials that allow students to manipulate materials with their hands to learn quantity and measurement.

Table 6-1: Average treatment effect in year 1 (basic result)

	Model 2-(1)	**Model 2-(2)**	**Model 2-(3)**	**Model 2-(4)**	**Model 2-(5)**	**Model 2-(6)**
(Intercept)	0.013	0.102	0.100	0.306	0.457	0.476
	(0.088)	(0.089)	(0.089)	(0.663)	(0.669)	(0.674)
Treatment	0.588^{**}	0.533^{**}	0.544^{***}	0.526^{***}	0.475^{***}	0.486^{**}
	(0.080)	(0.082)	(0.081)	(0.072)	(0.072)	(0.071)
La Union	-0.193	-0.122	-0.136	-0.061	-0.083	-0.103
	(0.141)	(0.116)	(0.113)	(0.190)	(0.157)	(0.154)
San Miguel	-0.273^{**}	-0.204^{*}	-0.210^{**}	-0.111	-0.149	-0.163
	(0.113)	(0.106)	(0.106)	(0.151)	(0.154)	(0.153)
San Vicente	-0.033	-0.109	-0.100	-0.055	-0.162	-0.168
	(0.127)	(0.121)	(0.121)	(0.138)	(0.137)	(0.135)
Urban	0.324	-0.022	-0.016	0.010	-0.188	-0.189
	(0.200)	(0.235)	(0.239)	(0.224)	(0.222)	(0.223)
La Union×Urban	-0.184	0.062	0.072	0.189	0.572^{**}	0.576^{**}
	(0.270)	(0.277)	(0.280)	(0.295)	(0.279)	(0.280)
San Miguel×Urban	-0.073	0.112	0.115	-0.080	0.260	0.268
	(0.241)	(0.264)	(0.268)	(0.225)	(0.240)	(0.240)
San Vicente×Urban	-0.095	0.040	0.034	0.296	0.415^{*}	0.418^{*}
	(0.258)	(0.290)	(0.293)	(0.253)	(0.250)	(0.249)
Z score baseline		0.511^{***}			0.517^{***}	
		(0.042)			(0.031)	
Z score baseline (2)			0.503^{***}			0.507^{**}
			(0.042)			(0.030)
Student characteristics	No	No	No	Yes	Yes	Yes
Teacher characteristics	No	No	No	Yes	Yes	Yes
School characteristics	No	No	No	Yes	Yes	Yes
R^2	0.093	0.279	0.276	0.205	0.370	0.367
Adj. R^2	0.090	0.277	0.274	0.184	0.353	0.349
N. obs.	3,032	3,032	3,032	3,030	3,030	3,030
N. Clusters	235	235	235	235	235	235

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data sources are baseline and end-line surveys of this research.

(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.

(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.

(4) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

Table 6-2: Average treatment effect in year 1 by cognitive skills of test items

	Knowin (1)	Knowin (2)	Applyin (1)	Applyin (2)	Reasonin (1)	Reasonin (2)
(Intercept)	0.364	0.504	0.260	0.424	-0.271	-0.166
	(0.618)	(0.629)	(0.594)	(0.598)	(0.594)	(0.603)
Treatment	0.553***	0.520***	0.335***	0.297***	0.243***	0.218***
	(0.068)	(0.068)	(0.066)	(0.064)	(0.065)	(0.064)
Z score baseline (2)		0.416***		0.487***		0.314***
		(0.027)		(0.029)		(0.031)
R^2	0.211	0.32	0.13	0.28	0.09	0.15
Adj. R^2	0.18	0.30	0.10	0.26	0.06	0.12
N. obs.	3,03	3,03	3,03	3,03	3,03	3,03
N Clusters	235	235	235	235	235	235

***p < 0.01; **p < 0.05; *p < 0.1

(1) Data sources are baseline and end-line surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted byfinite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(4) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

Table 6-3: Average treatment effect in year 1 by cognitive domains]

	Number & Operations (1)	Number & Operations (2)	Quantity & Measure (1)	Quantity & Measure (2)
(Intercept)	0.290	0.457	0.325	0.440
	(0.692)	(0.717)	(0.524)	(0.499)
Treatment	0.574***	0.535***	0.123*	0.096
	(0.071)	(0.069)	(0.066)	(0.066)
Z score baseline (2)		0.498***		0.343***
		(0.030)		(0.024)
R^2	0.201	0.353	0.118	0.208
Adj. R^2	0.179	0.335	0.094	0.186
N. obs.	3,030	3,030	3,030	3,030
N Clusters	235	235	235	235

***p < 0.01; **p < 0.05; *p < 0.1

(1) Data sources are baseline and end-line surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(4) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

Thus, the results from the basic specification demonstrate that the package of interventions improved student math learning in year 1. However, the absolute level of achievement shows the challenges for the ESMATE programme. At the end-line survey, 27.1 percent of the students in the treatment group could not correctly respond to the two-digit addition item ("35+21"), even after the one-year intervention. Around half of students correctly answered the two-digit addition item but could not solve the three-digit addition item ("253+174"). Students in the treatment group also had a difficulty in solving problems posed in texts. At the end-line survey, 8.0 percent of students in the treatment group correctly answered the three-digit addition item but could not solve the problem posed in texts that involved a two-digit addition with carrying (compare the percentage with the correct response rate to the three-digit addition item, "253+174", which was 28.3 percent in the treatment group). The ESMATE programme improved student math learning, but the absolute level of achievement demonstrates the need to continue to improve the package in the future.

This research collected process data through interviews with students and teachers, and math lesson observations. While the percentage of students who had a math textbook (other than the ESMATE textbook) in the control group was around 20 percent, almost all the teachers in the group prepared math lesson plans regularly, referring to either existing math textbook, teacher's guidebook etc. They provided math instructions according to the materials. Teachers in the treatment group referred to the ESMATE teacher's guidebook in their preparation of math classes. The percentage of teachers who prepared a plan for what to present on the black board was larger in the treatment group (64.1 percent) than the control group (15.3 percent).

In order to facilitate math learning through problem solving in class, each page of the ESMATE textbook, which corresponds to a class, is structured along four steps: (1) show the theme of the lesson; (b) pose problem examples; (c) explain the general principle, and (d) provide exercises.[21] During math lessons, the percentage of teachers who walk around

[21] Please refer to the appendix of this paper for more the detail.

classroom to check student notebooks was higher in the treatment group (90.6 percent) than the control group (79.7 percent). Teachers in the treatment group more frequently instructed students to try again if they answered a question incorrectly than the control group did (70.9 percent in the treatment group, and 58.5 percent in the control group). The percentage of teachers who assigned math homework four or more times in a week was larger in the treatment group (81 percent) than the control group (44 percent), which led students in the treatment group to better reinforce their math learning at home more often than the control group. The ESMATE textbook facilitated math learning through problem solving in class and at home.

The impact estimates discussed so far show the effect of the intervention package as a whole. To shed light on the mechanism underlying it, we attempted further regressions focusing on the availability of textbooks at schools for students' use. Before the intervention with ESMATE textbook distribution, some schools already had a stock of existing textbooks (other than ESMATE textbooks) to be used by students. As noted in Table 3-1, around 30 percent of the students respectively in the treatment and control groups used existing textbooks in a math class in 2017, the school year prior to the start of the intervention for our impact evaluation.[22] In these cases, most schools kept existing textbooks and lent them to students. Since the number of existing textbooks was not sufficient, two or more students shared a textbook. There were around 90 schools in each of the treatment and control groups, in which students learned math using (and usually sharing) existing textbooks. In order to investigate the average impact of the ESMATE programme's distribution of new textbooks, we conducted regression analysis using the sub-sample of schools. The estimated impact on the sub-sample, shown in Table 7, does not differ much from the estimated value of the overall sample.

[22] Almost all of the teachers planned lessons referring to either the textbook, teachers' guide, reference book, or curriculum.

Table 7: Average treatment effect in year 1 focusing on the effect of ESMATE textbook (using the sub-sample of schools where textbook was available in the school year previous to the interventions)

	Model 2-(7)	**Model 2-(8)**	**Model 2-(9)**
(Intercept)	0.218	0.439	0.455
	(0.703)	(0.759)	(0.767)
Treatment	0.541^{***}	0.492^{***}	0.503^{***}
	(0.083)	(0.083)	(0.082)
Z score baseline		0.503^{***}	
		(0.033)	
Z score baseline (2)			0.492^{***}
			(0.031)
R^2	0.220	0.381	0.378
Adj. R^2	0.195	0.360	0.358
N. obs.	2,520	2,520	2,520
N. Clusters	185	185	185

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data sources are baseline and end-line surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values areadjusted by finite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(4) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but notshown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

(2) The heterogeneity of one-year impacts of a package of interventions

The heterogeneity of impacts by the baseline score and the economic status is analyzed in two ways: applying equation (2) to sub-samples divided by each characteristic, and applying equations (3)-(4) with interaction terms to full samples. The results from the first approach are shown in Tables 8-1 and 8-2. The impact of treatment on student math learning becomes larger among higher quartiles defined on the baseline test scores. On the other hand, the impact does not show a monotonic relation with the household economic status.

Table 8-1: Heterogeneity of impact by baseline scores (sub-sample analysis)

	Q1	Q2	Q3	Q4
(Intercept)	-1.083	-0.772	1.561^{*}	2.600^{***}
	(0.939)	(0.894)	(0.851)	(0.953)
Treatment	0.368^{***}	0.451^{***}	0.453^{***}	0.512^{***}
	(0.083)	(0.098)	(0.082)	(0.109)
Z score baseline (2)	0.462^{**}	0.563^{**}	0.473^{***}	0.284^{***}
	(0.213)	(0.217)	(0.120)	(0.073)
R^2	0.243	0.254	0.288	0.400
Adj. R^2	0.163	0.165	0.214	0.302
N. obs.	845	763	850	572
N. Clusters	201	203	198	144

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Data sources are baseline and end-line surveys of this research.
(1) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(2) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(3) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(4) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

Table 8-2: Heterogeneity of impact by household economic status (sub-sample analysis)

	N. of Asset Types: 0-2	N. of Asset Types: 3	N. of Asset Types: 4	N. of Asset Types: 5	N. of Asset Types: 6-7
(Intercept)	-2.099	0.654	1.413^{*}	-0.781	1.600
	(1.830)	(0.832)	(0.823)	(0.958)	(1.055)
Treatment	0.455^{***}	0.571^{***}	0.474^{***}	0.362^{***}	0.519^{***}
	(0.163)	(0.109)	(0.089)	(0.105)	(0.097)
Z score baseline (2)	0.476^{***}	0.486^{***}	0.530^{***}	0.575^{***}	0.424^{***}
	(0.081)	(0.071)	(0.042)	(0.044)	(0.037)
N. of Asset types	-0.004				0.073
	(0.099)				(0.063)
R^2	0.469	0.488	0.409	0.432	0.408
Adj. R^2	0.302	0.402	0.346	0.366	0.343
N. obs.	307	501	768	702	752
N Clusters	150	189	214	193	177

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Data sources are baseline and end-line surveys of this research.
(1) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(2) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(3) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(4) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

The heterogeneity of impacts estimated using equations (3), (4) and (5) are shown in Table 8-3. Even after controlling for the total number of different types of student household assets in equation (4), the coefficient of the interaction term of the treatment assignment and baseline score is estimated to be 0.1, statistically significant at the 10 percent level. The result shows that the heterogeneous impact by the baseline score is not brought via the correlation with student economic status. The coefficient of the triple interaction term of the treatment assignment, baseline score level, and economic status is negative and statistically significant at the 5 percent level, which indicates that the heterogeneous impact by the baseline score level is larger for students with lower economic status.

Table 8-3: Heterogeneity of impacts by baseline score and household economic status

	Model 3-(1A)	**Model 3-(1B)**	**Model 4**	**Model 5**
(Intercept)	0.517	0.413	0.534	0.540
	(0.658)	(0.667)	(0.659)	(0.660)
Treatment	0.485^{***}	0.483^{***}	0.483^{***}	0.490^{***}
	(0.072)	(0.072)	(0.072)	(0.072)
Treatment×N. of Asset types	0.001		-0.009	-0.009
	(0.027)		(0.027)	(0.027)
Treatment×Z score baseline (2)		0.100^{*}	0.103^{*}	0.111^{**}
		(0.055)	(0.055)	(0.056)
Treatment×Z score baseline (2) ×N. of Asset types				-0.041^{**}
				(0.017)
Z score baseline (2)	0.505^{***}	0.454^{***}	0.453^{***}	0.452^{***}
	(0.030)	(0.045)	(0.045)	(0.045)
N. of Asset types (centered at zero)	0.026		0.032	0.032
	(0.019)		(0.020)	(0.020)
N. of Asset types		0.027^{**}		
		(0.013)		
R^2	0.363	0.365	0.365	0.367
Adj. R^2	0.347	0.349	0.349	0.350
N. obs.	3030	3030	3030	3030
N. Clusters	235	235	235	235

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data sources are baseline and end-line surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(4) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

The Early Grade Reading Assessment (EGRA) by USAID evaluated the student proficiency of letter and word identification (sound and letter, and sound and word) and reading fluency of the 2nd and 3rd grade students in El Salvador (Castro et al. 2018). According to their result, those students had difficulty identifying sounds and letters, and decoding. Even though the ESMATE textbook had a variety of graphics, they have to read Spanish text written in the textbook to understand the content. The students with a higher baseline score could read Spanish written in the textbook better than other students. Those students were also more prepared to learn by reading books. On the other hand, those students with a lower baseline score would have to catch up on reading Spanish, and also took time to acquire the skills necessary to learn by reading books.

(3) The accumulated impact of the first-year interventions in the following year

The accumulated impact on math learning in the following year of the first-year intervention is analyzed by applying equation (2) to the t=0 and t=2 panel data. In the training intervention by the ESMATE project, school principals in the treatment group were advised to continuously assign the same teacher for two years to the surveyed students. Although a similar direction was given to principals in control schools, the advice was better followed by principals in treatment schools than those in control schools.[23] The rate of assignment of the same teacher to the surveyed students in the treatment group was 76.6 percent in the year 2, which was higher than that in the control group (62.4 percent). Thus, teacher assignment in the year 2 and the treatment assignment of school are correlated with each other, which attenuates the estimated impact by simply controlling teacher characteristics in both years. Thus, the teacher characteristics of year 1 are controlled at first, and then the result is compared with the estimated value controlling teacher characteristics of year 2 for the robustness check. The accumulated impact of the whole

[23] Both the treatment and control groups were asked to assign the same teacher to the surveyed students for two years through a letter from the ministry.

package is estimated around 0.12 standard deviation of test scores, which is statistically significant at the 5 to 10 percent level (Table 9-1).[24] The parameter estimates are similar regardless of which teacher characteristics are controlled.

In 2019 (year 2), though students in both the treatment and control groups received the interventions, the difference in student's learning outcomes between those groups did not vanish. Kernel density curves (Figure 4) show that the distribution of Z scores between treatment and control groups gets closer than the end-line survey but the difference still remains.

[24] Considering student data attrition, we also conducted regression analysis of the impact with weighted sample. The sample was weighted with inverse probability weighting. The alternative result was almost the same as the values reported in Table 9-1. The estimated Lee bound of the impact is from 0.10 standard deviation (standard error: 0.06) to 0.15 standard deviation (standard error: 0.06).

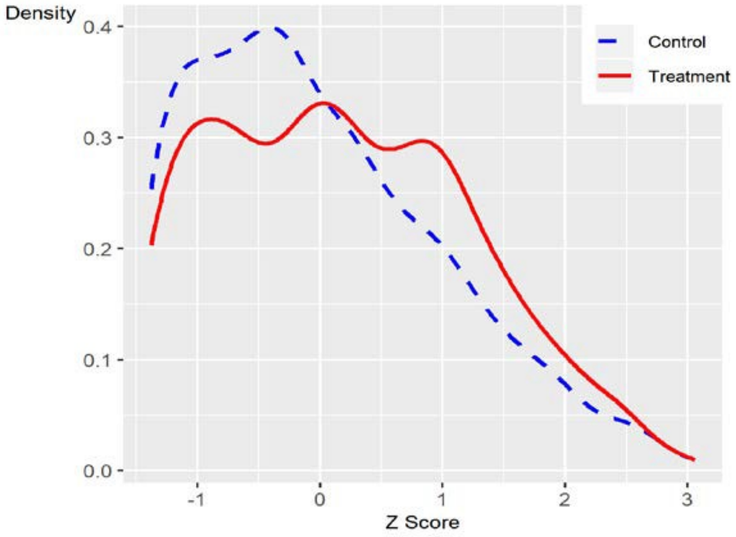
Table 9-1: Average treatment effect in year 2

	Model 2-(10)	Model 2-(11)	Model 2-(12)	Model 2-(13)	Model 2-(14)	Model 2-(15)	Model 2-(16)
(Intercept)	-0.056	-0.098	0.377	0.412	0.248	0.583	0.636
	(0.103)	(0.584)	(0.539)	(0.537)	(0.599)	(0.552)	(0.548)
Treatment	0.228***	0.183***	0.131**	0.140**	0.172**	0.112*	0.123**
	(0.077)	(0.066)	(0.063)	(0.063)	(0.069)	(0.063)	(0.062)
La Union	-0.118	0.063	0.108	0.087	-0.066	-0.041	-0.063
	(0.156)	(0.194)	(0.166)	(0.166)	(0.187)	(0.162)	(0.159)
San Miguel	-0.105	0.026	0.003	-0.014	0.042	-0.022	-0.033
	(0.133)	(0.154)	(0.141)	(0.140)	(0.146)	(0.135)	(0.134)
San Vicente	-0.145	-0.124	-0.191	-0.196	-0.183	-0.262**	-0.269**
	(0.126)	(0.137)	(0.129)	(0.128)	(0.129)	(0.130)	(0.128)
Urban	0.323**	0.309	0.088	0.087	0.114	-0.066	-0.078
	(0.150)	(0.188)	(0.194)	(0.192)	(0.176)	(0.177)	(0.173)
La Union×Urban	-0.258	-0.344	0.001	0.002	-0.259	0.133	0.148
	(0.250)	(0.293)	(0.262)	(0.262)	(0.273)	(0.227)	(0.225)
San Miguel×Urban	-0.089	-0.317	0.078	0.082	-0.197	0.191	0.193
	(0.205)	(0.228)	(0.225)	(0.226)	(0.222)	(0.210)	(0.209)
San Vicente×Urban	0.231	0.237	0.349*	0.354*	0.373*	0.467**	0.478**
	(0.225)	(0.223)	(0.202)	(0.199)	(0.201)	(0.189)	(0.185)
Z score baseline			0.518***			0.529***	
			(0.028)			(0.027)	
Z score baseline (2)				0.506***			0.517***
				(0.028)			(0.027)
Student	No	Yes	Yes	Yes	Yes	Yes	Yes
Teacher characteristics	No	Yes	Yes	Yes	No	No	No
Teacher characteristics	No	No	No	No	Yes	Yes	Yes
School	No	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.046	0.183	0.374	0.370	0.183	0.381	0.377
Adj. R^2	0.043	0.158	0.355	0.351	0.157	0.361	0.357
N. obs.	2,878	2,875	2,875	2,875	2,875	2,875	2,875
N. Clusters	235	235	235	235	235	235	235

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

(1) Data sources are baseline and follow-up surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted byfinite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores ateach round of survey.
(4) Student, teacher and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

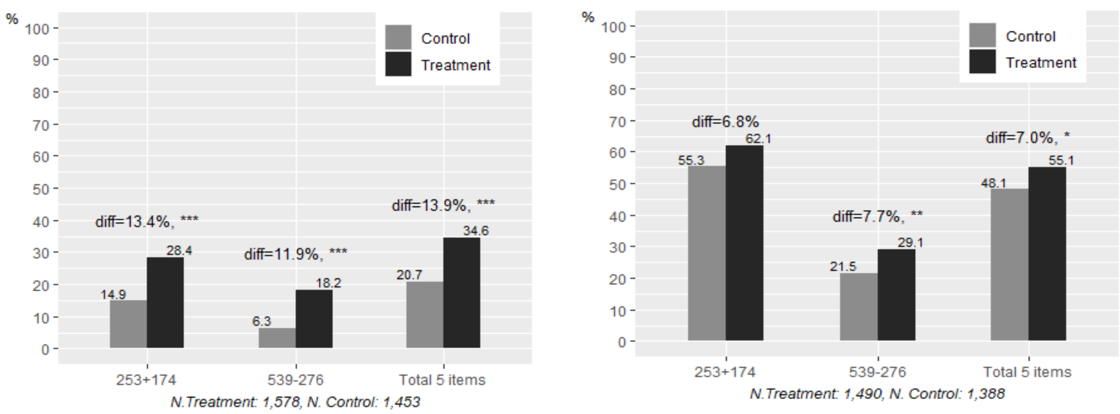
Figure 4: Kernel density curves of Z score at the follow-up



Note: Data source is follow-up survey of this research.

In year 2, the students in the control group caught up with those in the treatment group with respect to math learning thanks to the package of interventions including the ESMATE textbook. As explained in the previous section, the math test given at the follow-up survey included 5 questions that were exactly the same as those in the end-line survey assessment. Figure 5 presents the average correct response rate to the same items at the end of year 1 and year 2. At the end of year 1, more students in the treatment group correctly answered the item "539-276" than control groups by 11.9 percentage points. In year 2, the difference in average correct response rates to the same item decreased but still remained at 7.7 percentage points.

Figure 5: Correct response rates to the same math items at the end-line and follow-up surveys (Left: End-line and Right: Follow-up)



Note: Data sources are end-line and follow-up surveys of this research.

To analyze the accumulated impact on student math learning in more detail, the test items at the follow-up survey are divided into several categories according to the math content: (a) math content learned in grade 2 (Q1 to Q5); (b) math content learned in grade 3 using knowledge and skills introduced in grade 2 (Q6, 7, 8, 9, 11 and 18); (c) new math content learned in grade 3 (Q10, 12, 13, 14, 15, 16, 17 and 19) that assesses skills of knowing; and (d) math content learned in grade 3 that assess skills of applying or reasoning (Q20 to Q25). For example, students learn three-digit numbers in the 2nd grade, and they learn four-digits in the 3rd grade (Q6 of the follow-up survey test). Students learn division in the 3rd grade (Q12 of the follow-up survey test). The category (b)-1 includes only the items on numbers and operations in category (b). The regression results of test scores by those different categories are shown in Table 9-2. The accumulated impacts on (a), (b)-1 and (c) are positive and statistically significant. The intervention in year 1 improved student math learning in the treatment group, then based on the improved learning, the students in the group could learn new content in grade 3 better than those in the control group.

Table 9-2: Average treatment effect in year 2 (accumulated impact by math content)

	Category (a)	Category (b)	Category (b)-1	Category (c)	Category (d)
(Intercept)	0.328	0.340	0.453	0.947*	0.317
	(0.558)	(0.517)	(0.541)	(0.559)	(0.466)
Treatment	0.136***	0.080	0.122**	0.147**	-0.018
	(0.051)	(0.060)	(0.056)	(0.064)	(0.067)
Z score baseline (2)	0.409***	0.449***	0.381***	0.440***	0.366***
	(0.024)	(0.026)	(0.025)	(0.027)	(0.031)
R^2	0.272	0.291	0.223	0.306	0.205
Adj. R^2	0.249	0.268	0.198	0.283	0.179
N. obs.	2,875	2,875	2,875	2,875	2,875
N. Clusters	235	235	235	235	235

***p < 0.01; **p < 0.05; *p < 0.1

(1) Data sources are baseline and follow-up surveys of this research.
(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(4) Student, teacher (2019) and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(5) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

(4) The additional impact during the 2nd year of interventions

Both the one-year impact in subsection (1) above and the accumulated impact in subsection (2) above are positive and statistically significant, but the difference between the treatment and control groups became smaller in year 2 as the students in the control group also received the treatment in year 2. It is possible to identify econometrically how much the control group students caught up learning in year 2 using a two-step estimation employed by Das et al. (2013). We first estimate

$$Y_{ijk2}=\alpha_2+\gamma_{2j}Y_{ijk1}+C_{ijk}\beta_{2c}+P_{mjk2}\beta_{2p}+S_k\beta_{2s}+D_k\beta_{2D}+\varepsilon_{ijk2} \quad (6)$$

using the subset of the control group students, where P_{mjk2} is a vector of characteristics of the school principle and teacher m of *j* grade at school *k* such as age, gender, educational qualification in 2019. From this regression, we can calculate the effect of the lagged dependent variable for both treatment and control groups (it is a predicted value for the control group and it is an estimated value as counterfactual for the treatment group). We subtract that effect from the observed value of the year 2 test scores and use it as the dependent variable. Thus we estimate

$$Y_{ijk2} - \hat{\gamma}_{2j} Y_{ijk1} = \alpha_2 + \delta_{(2\text{-}1)}\, Treatment_k + C_{ijk}\beta_{2c} + P_{mjk2}\beta_{2p} + S_k\beta_{2s} + D_k\beta_{2D} + \varepsilon_{ijk2} \quad (7)$$

using both control and treatment groups, where $\hat{\gamma}_{2j}$ is the fitted parameter estimated from regression (6). $Y_{ijk2} - \hat{\gamma}_{2j} Y_{ijk1}$ represents the difference of test scores between follow-up survey and end-line of child *i* in *j* grade in school *k*. If $\delta_{(2\text{-}1)}$ in (6) is negative, it shows that the scale of improvement of math learning in the control group is larger than that of the treatment group in year 2, confirming the catch-up effect.

The results are presented in Table 10-1 and 10-2. The estimated value of $\delta_{(2\text{-}1)}$ is -0.12, which shows that the 2nd year treatment helped the students in the control group of catching up math learning.

Table 10-1: Additional impact in year 2 (1st stage)

	Model 6
(Intercept)	1.125
	(0.878)
Z score end-line	0.539***
	(0.024)
La Union	-0.568**
	(0.276)
San Miguel	0.025
	(0.170)
San Vicente	-0.190
	(0.166)
Urban	-0.281
	(0.244)
La Union×Urban	-0.065
	(0.332)
San Miguel×Urban	0.169
	(0.288)
San Vicente×Urban	0.924***
	(0.264)
R^2	0.496
Adj. R^2	0.458
N. obs.	1,204
N. Clusters	118

***p < 0.01; **p < 0.05; *p < 0.1

(1) Data sources are baseline and end-line surveys of this research.

(2) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.

(3) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.

(4) Student, teacher (2019) and school characteristics and stratification variables are controlled in all regression but not shown.

Table 10-2: Additional impact in year 2 (2nd stage)

	Model 7
(Intercept)	0.694
	(0.546)
Treatment	-0.115^{*}
	(0.066)
La Union	0.060
	(0.145)
San Miguel	0.202
	(0.138)
San Vicente	-0.020
	(0.108)
Urban	0.048
	(0.158)
La nion×Urban	-0.238
	(0.229)
San Miguel×Urban	-0.152
	(0.203)
San Vicente×Urban	0.244
	(0.171)
R^2	0.126
Adj. R^2	0.095
N. obs.	2,503
N Clusters	235

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Data sources are baseline, end-line and follow-up surveys of this research.

(1) Robust standard errors are clustered at school level, and are in parenthesis. The values are adjusted by finite sample correction.
(2) Student test scores are standardized by mean and standard deviation of control group test scores at each round of survey.
(3) Student, teacher (2019) and school characteristics and stratification variables (department dummies, urban dummy, and the interactions) are controlled in all regression but not shown.
(4) Z score baseline (2) is baseline score which excludes question No. 2 and No. 4.

4. Discussion

Glewwe et al. (2009) conducted a randomized controlled trial in Kenya to estimate the impact of distribution of textbooks on learning outcomes of students in grades 3 to 8 in English, math, and science. Their experiment was continued for two years. The average one-year and two-year impacts of the treatment were close to zero, and not statistically significant. In contrast, we investigated the effectiveness of the package of interventions including the distribution of math textbooks in El Salvador, and confirmed that the ESMATE programme improved student math learning on average. Contexts of student learning in Kenya and El Salvador are different. In Kenya, the language of instruction and mother tongue are different. On the other hand, in El Salvador, both of them are Spanish, thus the language barrier to learning would be lower for students in El Salvador than Kenya. The contents of the intervention are also different: in El Salvador there was a combination of different types of interventions, including the development of the math textbook. The results of the experiment in Kenya suggested that the contents of the national government textbook distributed were too hard for most students (Glewwe et al. 2009). In spite of those differences, the result of this study showed that the distribution of textbooks which were carefully designed considering teaching and learning in the country could improve student math learning. The ESMATE textbooks improved student math learning in year 1, and helped students to advance learning from the content learned to new content in year 2.

This study identified the heterogeneous impact by student baseline scores. The students with higher baseline scores improved their test scores by a larger amount, after the interventions. Glewwe et al. (2009) found that the impact of distribution of textbooks on student learning was positive and statistically significant only for students with highest quintile scores. The result of this study is in line with their research in that the impact of the distribution of textbooks is larger on students with higher baseline scores. Textbooks are developed according to the national curriculum in the country. However, in developing countries, there are always students who do not reach to the level supposed in the curriculum at the beginning of each grade. Such mismatch between curriculum and student learning level is not easily resolved only through an intervention related to the textbook.

5. Conclusion

Textbooks are an important intermediate that links curriculum and students, and teachers and students. Textbooks are used to make annual teaching plans by teachers, and they describe the content and methodology of teaching and learning. Textbooks are used for study at home by students. If the math contents are carefully sequenced in the textbook from grade 1 to 6 considering teaching and learning practices in the country, students can advance their math learning from content learned to new content. The Ministry of Education in El Salvador designed a package of interventions for schools, with technical support by JICA, to bring changes to teaching and learning practices to improve student math learning. The ESMATE programme was composed of (a) distribution of ESMATE textbooks, student workbooks and teachers' guidebooks, (b) introductory teacher training on the textbook, (c) regular class observation by the school principal, (d) review meetings among teachers based on the results of tests, and (e) introductory training of representatives of the parent association. The average one-year impact on student math learning is estimated around 0.5 standard deviation of test scores. The impact was larger on students with higher baseline scores. The average accumulated impact of the first-year intervention in the following year is 0.12 standard deviation. The package of interventions improved math learning of 2nd grade students. The impact persisted even after schools in the control group also received the package of interventions.

The ESMATE programme improved student math learning, but the absolute level of achievement was not yet satisfactory and the difference in achievement level among students remained tangible. Textbooks are developed according to the national curriculum in the country.

However, in developing countries, there are always students who do not reach the level supposed in the curriculum at the beginning of each grade. Teachers have difficulty in coping with students at different achievement levels in a class. Such mismatch between curriculum and student proficiency level is not easily resolved only by an intervention related to the textbook. It would be important to design complementary interventions to support those students who are not at the level supposed in the curriculum to catch up on learning. For example, if schools organize remedial activity on basic reading and writing of Spanish for 2nd grade students at the beginning of the school year, it would help them to learn math using the ESMATE textbook. The remedial activity by math proficiency level would also give students at the lower level a chance to catch up with other students on understanding and skills. Designing such complementary interventions and evaluating their effectiveness are left for further research.

Reference

Bashir, S., M. Lockheed, E. Ninan, and J.P. Tan. 2018. *Facing Forward: Schooling for Learning in Africa. Africa Development Forum Series*. Washington DC: World Bank.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Scotlandand M. Walton. 2017. “From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application.” *Journal of Economic Perspectives*, 31 (4), Fall 2017, 73-102.

Bhula, R., M. Mahoney, and K. Murphy. 2020. Conducting cost-effectiveness analysis (CEA). https://www.povertyactionlab.org/resource/conducting-cost-effectiveness-analysis-cea. (Access date: August 4, 2020)

Castro, V., J. Mejia, C. Cummiskey, J. Pressley, K. Betts, and A. Gove. (RTI International) 2018. *Proyecto Educación pára la Niñez y Juventud Evaluación de Lectura Inicial en El Salvador: Informe Final*. USAID.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararamanet 2013. “School Inputs, Household Substitution, and Test Scores.” *American Economic Journal: Applied Economics* 5(2) : 29–57.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch 2014. “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education.” In Glewwe, P (Ed.). *Education Policy in Developing Countries* (pp. 285-338). Chicago: University of Chicago Press.

Glewwe, P., M. Kremer, and S. Moulin. 2009. “Many Children left behind? Textbooks and Test Scores in Kenya.” *American Economic Journal: Applied Economics* 1 (1): 112-35.

Glewwe, P. (ed.) 2013.2013. *Education Policy in Developing Countries*. Chicago: University of Chicago Press.

Hanushek, E. A., and L. Woessmann. 2012. “Schooling, educational achievement, and the Latin American growth puzzle.” *Journal of Development Economics* 99 (2): 497-512.

Jamison, D. T., B. Searle, K. Galda, and S.P. Heyneman. 1981. “Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement.” *Journal of Educational Psychology* 73 (4): 556–67

Japan International Cooperation Agency. 2019. *Project Completion Report of Project for the Improvement of Mathematics Teaching in Primary and Secondary Education in El Salvador. (in Japanese).*

Koei Research & Consulting Inc. 2018. *Impact Evaluation for the Project for the Improvement of Mathematics Teaching in Primary and Secondary Education in El Salvador: Baseline Survey Report*. JICA. Unpublished.

———. 2019. *Impact Evaluation for the Project for the Improvement of Mathematics Teaching in Primary and Secondary Education in El Salvador: End-line Survey Report*. JICA. Unpublished.

Kremer, M., C. Brannen, and R. Glennerster. 2013. “The Challenge of Education and Learning in the Developing World.” *Science* 340: 297.

Kuecken, M. and M. A. Valfort. 2013. “When do textbooks matter for achievement? Evidence from African primary schools.” *Economics Letters* 119: 311–15.

McEwan, P.J. 2015. “Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments.” *Review of Educational Research* 85 (3): 353– 94.

Ministerio de Educatión, El Salvador. 2014. EDUCACIÓN DE EL SALVADOR EN CIFRAS 2009-2014.

———.2016. BOLETÍN ESTADÍSTICO N° 19.

———. 2016. Bases de Datos por Centros Escolares - Censo Escolar: Inicial Directorio 2016. El Salvador. Retrieved from https://www.mined.gob.sv/estadisticas-educativas/item/6116-bases-de-centros.html. (Access date: July 20, 2020).

———. 2018. Bases de Datos por Centros Escolares - Censo Escolar: Inicial Directorio 2018. El Salvador. Retrieved from https://www.mined.gob.sv/estadisticas-educativas/item/6116-bases-de-centros.html. (Access date: July 20, 2020).

———. 2018. Bases de Datos por Estudiante - Censo Escolar Inicial: Año 2018. Retrieved from https://www.mined.gob.sv/estadisticas-educativas/item/7154-bases-de-datos-spss.html. (Access date: July 20, 2020).

Mullis, I.V.S., M.O. Martin, and P. Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., and Galia, J.). 2008. *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Boston College.

UNDP 2017. *The Millennium Development Goals Report 2015*.

UNESCO Institute for Statistics (UIS). 2019. *UIS. Stat.* Retrieved from data.uis.unesco.org/ (Access date: February 14, 2020).

UNESCO. 2017. *More than One-half of Children and Adolescents Are Not Learning Worldwide. Fact Sheet No. 46* (UIS/FS/2017/ED/46).

USAID. 2017. *EVALUACIÓN DEL SECTOR EDUCACIÓN DE EL SALVADOR USAID/EL SALVADOR.*

Snilstvelt, B., J. Stevenson, R. Menon, D. Phillips, E. Gallagher, M. Geleen, H. Jobse, T. Schmidt, and E. Jimenez. 2016. *The Impact of education programmes on learning and school participation in low- and middle- income countries*. London: International Initiative for Impact Evaluation.

World Bank. 2001. *A Chance to Learn: Knowledge and Finance for Education in Sub-Saharan Africa.* Washington DC: World Bank.

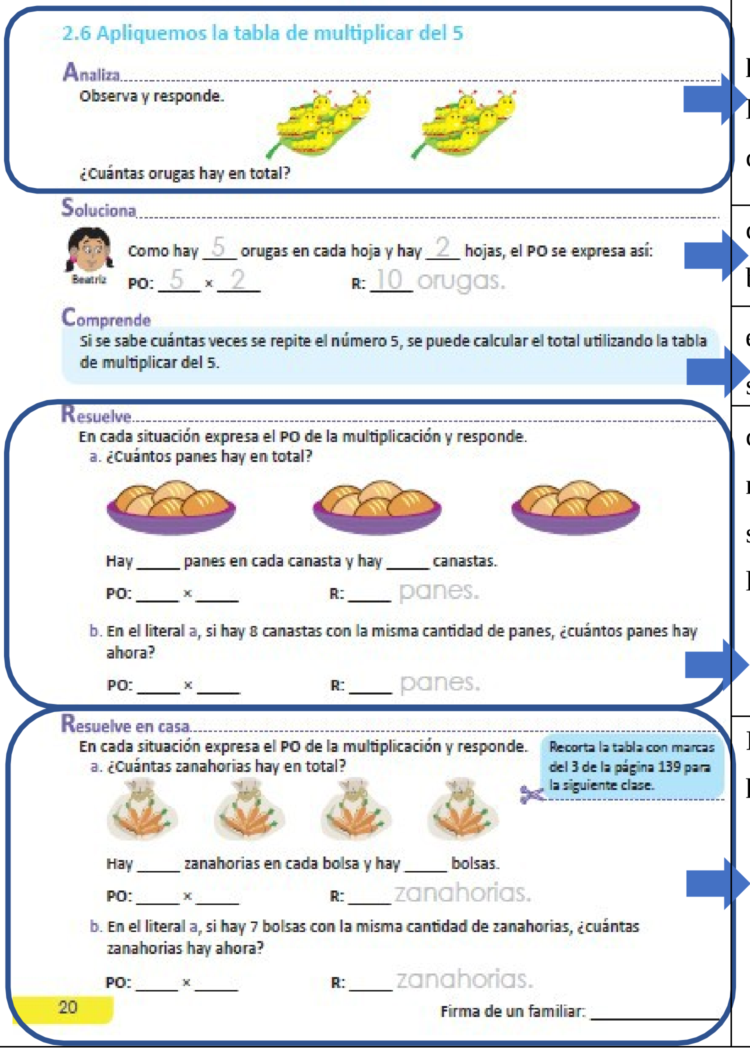
———. 2018. *World Development Report 2018: Learning to Realize Education's Promise.* Washington DC: World Bank.

———. 2020. *World development indicators.* Retrieved from https://data.worldbank.org/indicator/ (Access date: July 8, 2020).

Appendix 1: Sample pages of teaching and learning materials developed by the ESMATE project[1]

1. Textbook (primary 2nd grade)

The standard learning process structured in the ESMATE textbook is (1) students understand the theme of the lesson with a concrete problem example, (2) they work on the problem example with the help of the teacher and explanation in the textbook, (3) they understand the mechanism (general principle) behind the example written in the textbook, and (4) they work on exercises in the textbook with support from the teacher.

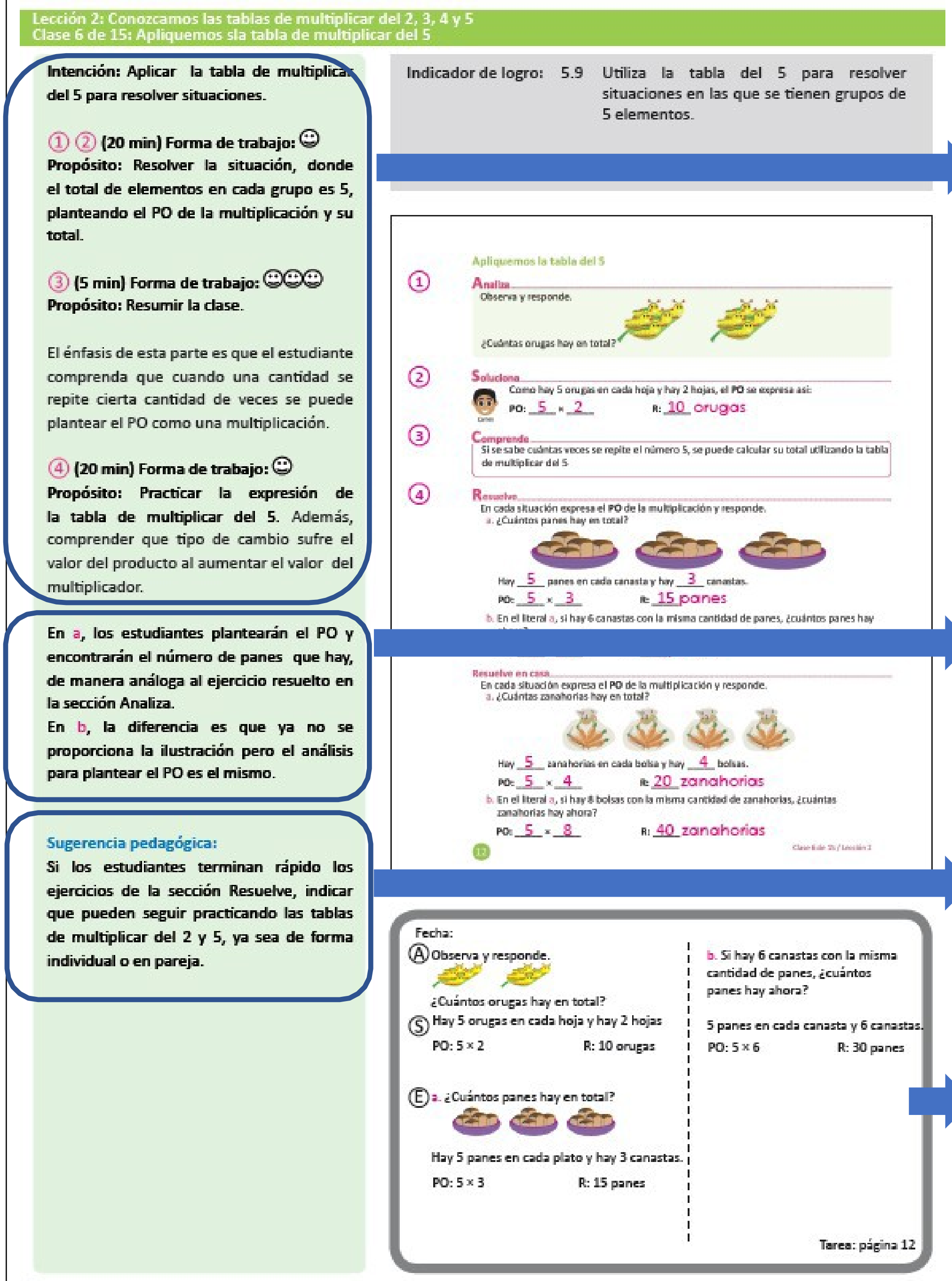
Learning process	Sample page	By each part of textbook, teacher	By each part of textbook, student
(1)	 2.6 Apliquemos la tabla de multiplicar del 5 Analiza Observa y responde. ¿Cuántas orugas hay en total?	pose the theme of lesson topic with concrete example	understand the theme of lesson topic
(2)	Soluciona Beatriz Como hay 5 orugas en cada hoja y hay 2 hojas, el PO se expresa así: PO: 5 × 2 R: 10 orugas.	check student note book	solve example problem
(3)	Comprende Si se sabe cuántas veces se repite el número 5, se puede calcular el total utilizando la tabla de multiplicar del 5.	explain how to solve	understand how to solve
(4)	Resuelve En cada situación expresa el PO de la multiplicación y responde. a. ¿Cuántos panes hay en total? Hay ____ panes en cada canasta y hay ____ canastas. PO: ____ × ____ R: ____ panes. b. En el literal a, si hay 8 canastas con la misma cantidad de panes, ¿cuántos panes hay ahora? PO: ____ × ____ R: ____ panes.	check student notebook and support student learning	practice through solving problem
	Resuelve en casa En cada situación expresa el PO de la multiplicación y responde. a. ¿Cuántas zanahorias hay en total? Recorta la tabla con marcas del 3 de la página 139 para la siguiente clase. Hay ____ zanahorias en cada bolsa y hay ____ bolsas. PO: ____ × ____ R: ____ zanahorias. b. En el literal a, si hay 7 bolsas con la misma cantidad de zanahorias, ¿cuántas zanahorias hay ahora? PO: ____ × ____ R: ____ zanahorias. 20 Firma de un familiar: ____	Instruct and check homework	practice through solving problem at home

Source: JICA (2019)

[1] Page 31 of JICA (2019) is translated in English and adapted by authors for this appendix.

2. Teacher's guidebook (primary 2nd grade)

Referring to the corresponding sample page of the textbook, technical guidance for teachers are provided in the guidebook. Teachers read the guidebook for their lesson preparation.

Sample page	Guidance in the guidebook for teachers
	— This part explains the purpose and time allocation of lesson — This part explains the intention of exercises. — This part suggests the additional exercise for students — This part shows board plan

Lección 2: Conozcamos las tablas de multiplicar del 2, 3, 4 y 5
Clase 6 de 15: Apliquemos sla tabla de multiplicar del 5

Intención: Aplicar la tabla de multiplicar del 5 para resolver situaciones.

① ② **(20 min) Forma de trabajo:**
Propósito: Resolver la situación, donde el total de elementos en cada grupo es 5, planteando el PO de la multiplicación y su total.

③ **(5 min) Forma de trabajo:**
Propósito: Resumir la clase.

El énfasis de esta parte es que el estudiante comprenda que cuando una cantidad se repite cierta cantidad de veces se puede plantear el PO como una multiplicación.

④ **(20 min) Forma de trabajo:**
Propósito: Practicar la expresión de la tabla de multiplicar del 5. Además, comprender que tipo de cambio sufre el valor del producto al aumentar el valor del multiplicador.

En a, los estudiantes plantearán el PO y encontrarán el número de panes que hay, de manera análoga al ejercicio resuelto en la sección Analiza.
En b, la diferencia es que ya no se proporciona la ilustración pero el análisis para plantear el PO es el mismo.

Sugerencia pedagógica:
Si los estudiantes terminan rápido los ejercicios de la sección Resuelve, indicar que pueden seguir practicando las tablas de multiplicar del 2 y 5, ya sea de forma individual o en pareja.

Indicador de logro: 5.9 Utiliza la tabla del 5 para resolver situaciones en las que se tienen grupos de 5 elementos.

Apliquemos la tabla del 5
① Analiza
Observa y responde.
¿Cuántas orugas hay en total?
② Soluciona
Como hay 5 orugas en cada hoja y hay 2 hojas, el PO se expresa así:
PO: 5 × 2 R: 10 orugas
③ Comprende
Si se sabe cuántas veces se repite el número 5, se puede calcular su total utilizando la tabla de multiplicar del 5.
④ Resuelve
En cada situación expresa el PO de la multiplicación y responde.
a. ¿Cuántos panes hay en total?
Hay 5 panes en cada canasta y hay 3 canastas.
PO: 5 × 3 R: 15 panes
b. En el literal a, si hay 6 canastas con la misma cantidad de panes, ¿cuántos panes hay ahora?

Resuelve en casa
En cada situación expresa el PO de la multiplicación y responde.
a. ¿Cuántas zanahorias hay en total?
Hay 5 zanahorias en cada bolsa y hay 4 bolsas.
PO: 5 × 4 R: 20 zanahorias
b. En el literal a, si hay 8 bolsas con la misma cantidad de zanahorias, ¿cuántas zanahorias hay ahora?
PO: 5 × 8 R: 40 zanahorias

Fecha:
Ⓐ Observa y responde.
¿Cuántas orugas hay en total?
Ⓢ Hay 5 orugas en cada hoja y hay 2 hojas
PO: 5 × 2 R: 10 orugas
Ⓔ a. ¿Cuántos panes hay en total?
Hay 5 panes en cada plato y hay 3 canastas.
PO: 5 × 3 R: 15 panes
b. Si hay 6 canastas con la misma cantidad de panes, ¿cuántos panes hay ahora?
5 panes en cada canasta y 6 canastas.
PO: 5 × 6 R: 30 panes
Tarea: página 12

Source: JICA (2019)

Appendix 2: Supplementary Tables

Table A-1: Composition of math item (Baseline)

Item No.	Item content	Cognitive domain	Cognitive skill	Answer format
1	To write the number of quantity represented by the image	NO	Knowing	Fill in blank
2	To add; one-digit number plus one-digit number = one-digit number	NO	Knowing	Fill in blank
3	To subtract; one-digit number minus one-digit number = one-digit number	NO	Knowing	Fill in blank
4	To find which side is longer and by how much using an eraser as unit to measure, in the notebook	QM	Knowing	Multiple choice
5	To fill out the number pattern of 2 by 2	NO	Knowing	Fill in blank
6	To choose the rectangle among different two dimensional figures	Geometry	Knowing	Multiple choice
7	To choose the watch that indicates a determined time	QM	Knowing	Multiple choice
8	To choose the container with more liquid through indirect comparison	QM	Knowing	Multiple choice
9	To add; two-digit number plus two-digit number = two-digit number without carrying	NO	Knowing	Fill in blank
10	To subtract; two-digit number minus two-digit number = two-digit number without borrowing	NO	Knowing	Fill in blank
11	To fill out the number pattern of 6 by 6	NO	Knowing	Fill in blank
12	To place numbers up to 50 on the number line	NO	Applying	Fill in blank
13	To solve the written problem of addition; one-digit plus one-digit = two-digit number	NO	Applying	Fill in blank
14	To solve the written problem of subtraction; two-digit minus one-digit = one-digit number	NO	Applying	Fill in blank
15	To solve the written problem of addition; add 7 times 5	NO	Applying	Fill in blank
16	To solve the problem of equivalence between the amount of value and a few coins	QM	Applying	Fill in blank
17	To find the unknown number in accumulation of the same number in an addition problem	NO	Applying	Fill in blank
18	To solve the problem of accumulation of the same number adding one more time	NO	Reasoning	Fill in blank
19	To solve the problem of which two operations required and the result of the first operation is a number of the second operation	NO	Reasoning	Fill in blank
20	To solve problem of addition in which one of the numbers is the ordinary number	NO	Reasoning	Fill in blank

Test items are developed by the ESMATE project in this research.
Note: NO: Number and Operation; QM: Quantity and Measurement

Table A-2: Composition of math item (End-line)

Item No.	Item content	Cognitive domain	Cognitive skill	Answer format
1	To write three-digit number 109 represented by an image	NO	Knowing	Fill in blank
2	To add with numbers up to two-digits, without carrying	NO	Knowing	Fill in blank
3	To subtract numbers up to two-digit without borrowing	NO	Knowing	Fill in blank
4	To measure an object in centimeters	QM	Knowing	Fill in blank
5	To multiply 2 times one number	NO	Knowing	Fill in blank
6	To choose the square among two dimensional figures	Geometry	Knowing	Multiple choice
7	To identify the time of an event with determined duration	QM	Knowing	Fill in blank
8	To read the capacity in liter	QM	Knowing	Fill in blank
9	To add three-digit numbers with carrying to the position of hundreds	NO	Knowing	Fill in blank
10	To subtract three-digit numbers with borrowing from the position of hundreds	NO	Knowing	Fill in blank
11	To multiply 6 times one number	NO	Knowing	Fill in blank
12	To place a three-digit number in the number line	NO	Applying	Fill in blank
13	To solve the written problem of addition with the numbers up to two-digits with carrying to the position of tens	NO	Applying	Fill in blank
14	To solve the written problem of the subtraction with the numbers up to two-digit with borrowing from the position of tens	NO	Applying	Fill in blank
15	To write the multiplication from given addition of the accumulation with the same numbers and find the answer	NO	Applying	Fill in blank
16	To solve the written problem of the situation of a purchase to find the combination of currencies to be given as the change	QM	Applying	Multiple choice
17	To solve written problem of ratio	NO	Applying	Fill in blank
18	To fill in blanks on both sides of a number sentence that uses multiplication and addition.	NO	Reasoning	Fill in blank
19	To solve the problem of subtraction as an inverse operation of addition	NO	Reasoning	Fill in blank
20	To solve an addition with three numbers, one of which is unknown	NO	Reasoning	Fill in blank

Test items are developed by the ESMATE project in this research.
Note: NO: Number and Operation; QM: Quantity and Measurement

Table A-3: Composition of math item (Follow-up)

Item No.	Item content	Cognitive domain	Cognitive skill	Answer format
1	To add three-digit numbers with carrying to the position of hundreds	NO	Knowing	Fill in blank
2	To subtract three-digit numbers with borrowing from the position of hundreds	NO	Knowing	Fill in blank
3	To multiply 6 times one number	NO	Knowing	Fill in blank
4	To solve the problem of the addition with the numbers up to two-digit with carrying to the position of tens	NO	Applying	Fill in blank
5	To solve the addition with three numbers, one of which is unknown	NQR	Reasoning	Fill in blank
6	To write four-digit number represented by an image	NO	Knowing	Fill in blank
7	To add with four-digit numbers, with carrying one time	NO	Knowing	Fill in blank
8	To subtract with four-digit numbers with borrowing one time	NO	Knowing	Fill in blank
9	To measure an object in meters and centimeters	QM	Knowing	Fill in blank
10	To multiply two-digit number and one-digit number without carrying	NO	Knowing	Fill in blank
11	To choose the rectangle among the different two dimensional figures	Geometry	Knowing	Multiple choice
12	To divide two-digit number by one-digit number without remainder	NO	Knowing	Fill in blank
13	To represent the capacity writing the fraction	NO	Knowing	Fill in blank
14	To multiply three-digit number and one-digit number with carrying one time	NO	Knowing	Fill in blank
15	To divide two-digit number by one-digit number with remainder	NO	Knowing	Fill in blank
16	To calculate the addition and subtraction combined with brackets	NO	Knowing	Fill in blank
17	To read the number line of fractions	NO	Knowing	Fill in blank
18	To interpret the bar chart	NQR	Knowing	Fill in blank
19	To calculate the diameter from the radius of a circle	Geometry	Applying	Fill in blank
20	To solve the written problem of the addition with four-digit numbers with carrying	NO	Applying	Fill in blank
21	To solve the written problem of multiplication with three-digit numbers by one-digit number	NO	Applying	Fill in blank
22	To solve the written problem of division with two-digit numbers by one-digit number	NO	Applying	Fill in blank
23	To find the length of one edge from the total length of three edges of an equilateral triangle	Geometry / QM	Applying	Fill in blank
24	To find a four-digit number, based on some hints using round up and off	NO	Reasoning	Fill in blank
25	To find the correct answer of written problem of division in which it is required add 1 to the quotient	NO	Reasoning	Fill in blank

Test items are developed by the ESMATE project in this research.

Note: NO: Number and Operation; QM: Quantity and Measurement; NR: Numerical and Quantitative elation

Table B-1: Characteristics of students (baseline) remained at the end-line survey

Content	Treatment	Control	Mean Diff.	Adjusted Mean Diff. (a)	Standard Error of (a)	P-Value of (a)
Morning Shift (%)	96.32	91.40	4.92	5.87	2.85	0.04 * *
Age	7.77	7.66	0.08	0.04	0.04	0.32
_sd	0.74	0.69				
Sex (Male) (%)	48.35	50.89	-2.54	-2.86	2.21	0.19
N. elder brother/sister	1.59	1.54	0.05	0.06	0.08	0.42
_sd	1.81	1.76				
N. younger brother/sister	0.81	0.82	0.00	-0.02	0.03	0.64
_sd	0.94	0.91				
Test score						
Raw test score (Total points: 20)	5.22	4.97	0.25	0.39	0.28	0.16
_sd	3.42	3.69				
Raw test score (Total points: 18, which excludes Q2&Q4)	3.87	3.71	0.15	0.28	0.24	0.25
_sd	3.02	3.24				
Asset of study						
Math textbook 2017 (%)	31.14	37.76	-6.62	-6.90	3.53	0.05 * *
Math notebook 2017 (%)	87.44	88.03	-0.59	-1.98	2.91	0.49
Notebook only for Math 2017 (%)	81.04	80.81	0.23	-1.35	3.16	0.67
Own Study Desk at Home (%)	32.19	33.70	-1.51	-0.92	2.55	0.71
Asset of student household						
Smartphone (%)	74.89	74.42	0.47	0.26	2.63	0.92
Computer (%)	25.68	23.80	1.89	1.50	1.94	0.44
Refrigerator (%)	83.77	82.26	1.51	1.43	1.82	0.43
Car (%)	32.85	31.43	1.42	1.04	2.28	0.64
TV (%)	90.81	91.95	-1.15	-1.10	1.32	0.4
Tap water (%)	79.33	81.02	-1.69	-1.44	2.43	0.55
Electricity (%)	95.56	94.91	0.65	0.47	1.38	0.73
Flush Toilet (%)	56.25	52.68	3.56	3.36	3.09	0.27
Using wood for cooking (%)	57.32	58.60	-1.27	-1.42	3.19	0.65
Using gas for cooking (%)	90.11	90.23	-0.13	-0.16	1.36	0.9
Using electricity for cooking (%)	6.28	4.61	1.67	0.60	1.64	0.71
N. of schools	117	121				
N. of students	1939	1846				

Data sources are baseline and end-line surveys of this research.
Note: (1) 10% significance: *, 5% significance: **, 1% significance: ***. Adjusted mean difference in this table is obtained by regressing the value of each characteristics on treatment assignment dummy with controlling stratification variables (department and urban/rural dummies, and the interactions). Robust standard errors are clustered at school level. (2) Binary values are Morning Shift (Morning shift: 1), Sex (Male:1), Textbook / Notebook (Yes:1, No:0), and Asset of study and student household (Yes:1, No:0).

Table B-2: Characteristics of students (baseline) remained at the follow-up survey

Content	Treatment	Control	Mean Diff.	Adjusted Mean Diff. (a)	Standard Error of (a)	P-Value of (a)
Morning Shift (%)	107.39	111.03	-3.64	4.60	3.00	0.12
Age	7.79	7.73	0.06	0.06	0.04	0.14
_sd	0.62	0.59				
Sex (Male) (%)	49.19	50.79	-1.60	-2.12	2.21	0.33
N. elder brother/sister	1.60	1.53	0.06	0.09	0.08	0.3
_sd	1.81	1.76				
N. younger brother/sister	0.81	0.81	0.00	0.00	0.04	0.97
_sd	0.94	0.90				
Test score						
Raw test score (Total points: 20)	5.27	4.93	0.34	0.47	0.28	0.09 *
_sd	3.43	3.66				
Raw test score (Total points: 18, which excludes Q2&Q4)	3.92	3.67	0.25	0.35	0.24	0.15
_sd	3.04	3.21				
Asset of study						
Math textbook 2017 (%)	31.07	37.18	-6.10	-6.78	3.66	0.06 *
Math notebook 2017 (%)	87.99	87.97	0.02	-1.57	2.89	0.58
Notebook only for Math 2017 (%)	81.21	81.56	-0.35	-2.28	3.08	0.45
Own Study Desk at Home (%)	31.81	33.79	-1.98	-1.43	2.50	0.56
Asset of student household						
Smartphone (%)	74.77	73.34	1.42	1.18	2.65	0.65
Computer (%)	25.37	23.13	2.24	1.44	1.90	0.44
Refrigerator (%)	84.36	82.20	2.16	2.03	1.81	0.26
Car (%)	33.09	31.27	1.82	1.16	2.21	0.59
TV (%)	91.01	91.50	-0.49	-0.71	1.35	0.59
Tap water (%)	80.07	80.98	-0.91	-1.08	2.45	0.65
Electricity (%)	95.44	94.60	0.84	0.69	1.43	0.63
Flush Toilet (%)	55.84	52.67	3.17	2.59	3.08	0.4
Using wood for cooking (%)	57.18	59.58	-2.40	-1.87	3.17	0.55
Using gas for cooking (%)	90.13	89.55	0.58	0.38	1.46	0.79
Using electricity for cooking (%)	6.44	4.11	2.34	1.33	1.59	0.4
N. of schools	117	118				
N. of students	1490	1388				

Data sources are baseline and end-line surveys of this research.
Note: (1) 10% significance: *, 5% significance: **, 1% significance: ***. Adjusted mean difference in this table is obtained by regressing the value of each characteristics on treatment assignment dummy with controlling stratification variables (department and urban/rural dummies, and the interactions). Robust standard errors are clustered at school level. (2) Binary values are Morning Shift (Morning shift: 1), Sex (Male:1), Textbook / Notebook (Yes:1, No:0), and Asset of study and student household (Yes:1, No:0).

Abstract (in Japanese)

要約

1990 年代以降、開発途上国における初等教育就学率は急速に向上した一方、教育の質については十分な改善が達成されていない。教授・学習の実践において教科書は、カリキュラム、教員及び生徒をつなぐ重要な教材である。教科書は、教授・学習の内容や方法を表すものであることから、教科書を改善することにより、教員の教授及び生徒の学習を向上しうる。本研究は、エルサルバドルにおいて、教科書配布等の介入パッケージにより生徒の算数の学習成果が向上するかを 2 年間にわたるランダム化比較試験により検証する。2 年生生徒の算数の学習成果に対し、1 年間の介入パッケージによる平均効果は算数テストスコアの 0.48 標準偏差と推定される。ベースラインスコアの高い生徒に対し、より大きな介入効果がみられた。実験開始から 2 年目には介入群・対照群の両方に教科書配布等が行われたが、1 年目の介入パッケージの 2 年目における平均累積効果は算数テストスコアの 0.12 標準偏差と推定される。実験 1 年目に介入パッケージは 2 年生生徒の算数の学習成果を向上させ、その効果は 2 年目に対照群が教科書配布等の介入パッケージを受けてからも持続した。

キーワード：教育開発、算数学習、算数教科書開発、人的資本、インパクト評価