HIAS-E-99

Does remedial activity with math workbook improve learning? Empirical evidence from scaled-up intervention in Niger

Takao Maruyama[(a)] and Takashi Kurosaki[(b)]

(a) Graduate School for Humanities and Social Sciences, Hiroshima University
(b) Institute of Economic Research, Hitotsubashi University

June, 2020

Hitotsubashi Institute for Advanced Study, Hitotsubashi University
2-1, Naka, Kunitachi, Tokyo 186-8601, Japan
tel:+81 42 580 8668 http://hias.hit-u.ac.jp/

Ver. June 30, 2020

Does remedial activity with math workbook improve learning? Empirical evidence from scaled-up intervention in Niger

Takao Maruyama and Takashi Kurosaki[1][2]

Abstract

A set of interventions, including distribution of math workbooks and training of activity facilitators, was scaled up by the government of Niger from 2017 to 2018, targeting approximately 310,000 students from 1st to 4th grade in around 3,500 public schools. The scaled-up interventions tried to help the students improve basic math learning through extra-curricular remedial activity. Because of budget constraint, the distribution of math workbooks was limited to students from 1st to 4th grade, not covering 5th and 6th grades. Focusing on the discontinuity of intervention between 4th and 5th grade students, this study investigated the impact of the interventions on student math learning using three-round survey data. The average impact of the interventions is estimated to be 0.36 to 0.38 standard deviations of test scores. The impact is larger for students with lower baseline scores.

Key words: Educational development; Basic math learning; Community participation; Scaling-up intervention; Sub-Saharan Africa

---

[1] Maruyama: Graduate School for Humanities and Social Sciences, Hiroshima University, 1-5-1, Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8529 Japan (e-mail: tmaru@hiroshima-u.ac.jp); Kurosaki: Institute of Economic Research, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan (e-mail: kurosaki@ier.hit-u.ac.jp).

[2] All data in this study are from JICA and were collected by the project "School for All: the Project on Support to Educational Development through Community Participation." We are grateful to Akiko Kageyama for her support for this study. We also appreciate Masahiro Hara and Nobuhiro Kunieda for their continuous encouragement. Findings, interpretations, and conclusions expressed in this paper are those of authors and do not represent the views of Japan International Cooperation Agency (JICA).

1. Introduction

Despite the progress in expansion of enrollment since 1990, approximately 617 million primary and lower secondary school-age children are not reaching minimum proficiency levels in reading and mathematics in the world (UNESCO, 2017). The learning crisis is particularly acute in Sub-Saharan Africa, where most children are not mastering basic reading and mathematics in primary education (Bashir et al., 2018). One of the reasons of the low-level learning is a mismatch between curricula and learning. Pritchett and Beatty (2015) reviewed basic reading assessments in India, Pakistan, Tanzania, Uganda, and Kenya and found that the learning progress per year of schooling was slow. Curricular expectations are too high, and the mismatch between curricular and children's learning levels expands with years of schooling in developing countries. Even though most children are left behind, teachers are required to complete prescribed syllabus, then their attention tends to be selective, often focusing on a limited number of children who can keep up (Banerji, 2000; Abadzi and Llambri, 2011).

Several systematic reviews in educational development agree that pedagogical intervention which tailors teaching to student learning level is effective at improving learning (Evans and Popova, 2016). An example that adapt the approach for improving basic reading and math is "Teaching at the Right Level (TaRL)" developed by Pratham, an Indian nongovernmental organization. TaRL divides 3rd to 5th grade students into several groups according to the assessment result. They are taught basic reading and math with various materials and activities that match with their proficiency levels. TaRL has helped children to improve basic reading and math proficiency rapidly (Banerjee et al., 2017). In the process of the scaled-up program by the government, strong leadership of state government and continuous on-site monitoring and support were essential elements for changing teaching practice successfully and improving learning (ibid). However, it is still challenging to scale up TaRL successfully in sub-Saharan Africa, where the capacities of education administration and teachers are often limited.

In the context of limited capacity of education administration and teachers, Niger has developed a package of interventions to improve basic math proficiency: “Paquet Minimum Axé sur la Qualité” (minimum package for quality learning or PMAQ) developed by the “Ecole pour tous” project (the EPT project) in cooperation with the Japan International Cooperation Agency (JICA). PMAQ seeks to improve basic math learning through extracurricular remedial activities by two types of interventions: a one-day training of School Management Committee (SMC) presidents and secretaries (school principals), and a set of distribution of math workbooks for each student and a two-day training of activity facilitators. PMAQ was introduced in pilot schools in two regions (Tillaberi and Niamey) of Niger’s eight regions, and students who participated extra-curricular remedial activity improved math learning in the 2013-14 school year (JICA, 2017).

Following the successful pilot, the government of Niger decided to scale up PMAQ to all public primary schools in Tillaberi region; however, because of budget constraint, the distribution of math workbooks was limited to 1st to 4th grades. Focusing on the discontinuity of intervention between 4th and 5th grade students, this study investigates the impact of PMAQ on student math learning. There are mainly three contributions of this paper to the literature. First, this study investigates the impact of scaled-up intervention on student math learning, using quasi-experimental data from three rounds of survey. While the number of studies on pilot interventions increases, rigorous evidence on the scaled-up intervention impact is still scarce in education. Although a small pilot succeeds, the scaled-up intervention by the government often faces implementation challenges (Banerjee et al., 2017; Bold et al., 2018). The ministry and the EPT project did not plan to conduct rigorous evaluation, but the discontinuity of intervention between 4th and 5th grades allowed us to investigate the impact of PMAQ. The average impact on student math learning is estimated to be 0.36 to 0.38 standard deviations. The pre-baseline and baseline data confirm that the identification assumption of the estimation is satisfied.

Second, this study evaluates the heterogeneity of impacts with respect to two aspects that are important in Niger. One of them is baseline score level. In Niger, 92.4 percent of students in last grade of primary education do not reach sufficient levels in mathematics, and

68.4 percent cannot solve basic math item such as four operations (PASEC, 2016). Girls' achievement level of mathematics is lower than boys (ibid). Thus, this study assesses the heterogeneous impact with respect to gender as well. It turns out that while the impact was larger on students with lower baseline scores, the impact additional for girls was not statistically significant. It demonstrates that the heterogeneous impact by baseline score level was not brought via the correlation with gender. PMAQ helped children left behind improve basic math learning.

Third, the results of this study indicate that PMAQ improved quality of remedial activity organized by SMC even in a context of limited capacity of educational administration and teachers. In Niger, district educational offices suffer from the lack of staff and logistical resources. One district educational office oversaw 249 SMCs on average (Kunieda et al., forthcoming). Less than one percent of primary school teachers possessed minimum knowledge of the subjects that they teach (World Bank, 2017). The minimum knowledge indicator reflects the percentage of teachers who scored above 80 percent in the lower primary portion of language and mathematics tests (ibid). In most sub-Saharan African countries, the capacity of teachers is limited like Niger (Bashir et al., 2018). Only 7 percent of teachers met the minimum content knowledge of language and mathematics in seven sub-Saharan African countries (Bold et al, 2017). PMAQ can help those countries with limited capacity of teachers rapidly overcome learning crisis.

The remainder of this paper is organized into the following 5 sections: (1) scaled-up intervention and timeline; (2) evaluation strategy; (3) results; (4) discussions and (5) conclusion.

2. Scaled-up intervention and timeline

(1) Scaled-up intervention

PMAQ seeks to improve basic math learning through extracurricular remedial activities organized by SMC[3]. For this purpose, two types of interventions were designed. The first intervention is a one-day training of SMC presidents and secretaries (school principals) that strengthen the capacity of SMC to develop school action plan integrating organization of remedial activity using math workbooks. After the training, school conducts basic math test to measure students' proficiency level. SMC organizes a community general assembly to share the result among teachers, parents and community, and discuss causes of low learning and potential activities[4]. Based upon the discussions, SMC prepares a draft of school action plan integrating remedial activity, and hold a community general assembly again to adopt it.

The second intervention of PMAQ is pedagogical, which consists of distribution of math workbooks and a two-day training of activity facilitators (teachers and community facilitators). The EPT project developed nine workbooks on introduction to number and basic four operations, which reflect different math proficiency levels[5]. Math workbooks which match with proficiency level are distributed for each student, according to the result of basic math assessment. Students learn basic math with workbooks by their own paces. Activity facilitators check their answers and give an instruction for each student. The content is arranged systematically so students can easily follow and progressively understand more mathematics

---

[3] For further detail of PMAQ, see Hara et al. (forthcoming) and Kunieda et al. (forthcoming).

[4] SMC facilitates the discussion at community general assembly using the analysis framework; namely (i) inadequate time spent on learning; (ii) lack of suitable learning materials and environments; and (iii) ineffective and low-quality teaching and learning practices. For each of these factors, the community identifies a corresponding solution (e.g., increasing the amount of time spent on learning, providing suitable learning materials and environments and improving the quality of teaching and learning practices). These solutions are then integrated into the SMC-organized extracurricular remedial activities.

[5] The math workbooks were issued in nine volumes:
Volume 0 (pre-math): Introduction to handwriting
Volume 1: Introduction to numbers: 0 – 10; Composition and decomposition of numbers
Volume 2: Introduction to numbers: 11-20; Addition/subtraction without carrying and borrowing; Addition and subtraction with carrying and borrowing
Volume 3: Addition and subtraction of 3-digit numbers; Understanding numbers: 21-79
Volume 4: Addition and subtraction of 2-digit numbers; Number sense 80 – 120; Addition and subtraction: vertical; Number sense: up to 1,000
Volume 5: Addition and subtraction of 2 to 3-digit numbers; Number sense: up to 10,000; Addition and subtraction of 2 to 3-digit numbers: vertical
Volume 6: Multiplication and division: basic; Division with remainders
Volume 7: Number sense: 10,000 – 100,000,000; Multiplication: vertical
Volume 8: Division: vertical; Calculation rules and techniques

principles. For example, students learn what a number is by following workbook pages that gradually transforms an illustration of a concrete object into a number. The math workbook includes plenty of well-developed math problems so students can practice their mathematics skills until they have mastery.

Out of Niger's eight regions, PMAQ was introduced in pilot schools in Tillaberi and Niamey regions. Since students who participated extra-curricular remedial activity improved math learning in the 2013-14 school year (JICA, 2017), the Government of Niger decided to scale up PMAQ in Tillaberi region in collaboration with multi-donor support[6]. The population of the region is around 2.6 million which accounts for about 16 percent in the country, and 53.8 percent live in poverty (national average: 59.5%). While the access to primary education in the region is close to the national average, the quality lags. 97.1 percent of students in the last year of primary school do not reach sufficient levels in mathematics in the region (PASEC, 2016). Gender gap is also an issue in educational development. In the region, primary gross enrollment ratio is 79.3 percent (boys) and 74.2 percent (girls), and primary completion rate is 75.3 percent (boys) and 67.2 percent (girls) (The Ministry of Primary Education in Niger, 2016).

As the scaled-up intervention, remedial activity was organized from March to June 2018 for the students in Tillaberi region; however, because of budget constraint, the distribution of math workbooks was limited to 1st to 4th grades of students.

(2) Timeline of scaled-up intervention

[Insert Figure 1: Timeline of scaled-up intervention]

The ministry originally scheduled the scale up of PMAQ in the 2016-17 school year[7]. From November to December 2016, a training of SMC president and secretary (school principle) on

[6] Multi-donor support was provided by the Global Partnership for Education, AFD, Swiss Cooperation and the World Bank.

[7] The Niger school year starts in October and ends in June the next year.

school action plan development was organized. Math workbooks were to have been delivered on March 2017; however, as the procurement process was significantly delayed, the ministry postponed the distribution of math workbooks and the training of activity facilitators to the following school year.

A two-day training of trainers on the use of math workbook was organized in February 2018 for activity facilitators who were selected by each SMC. In March, before the start of the remedial activity at SMCs, the Ministry organized a forum of stakeholders including departments of the Ministry of Primary Education, the director of local education administration office, inspectors, pedagogical advisors, executive members of SMC federation[8], SMC supervisors, and mayors. At the forum, the target of improvement in basic math learning was discussed, and the participants committed 30 percent point increase of correct response rate of 1st to 4th grade students through remedial activities. They also pledged to organize remedial activities for them at least 10 hours per week at each SMC[9].

According to the Ministry of Education in Niger (2018), 3,380 targeted schools planned and conducted remedial activities on basic math. Each SMC implemented an average of 133 hours of extracurricular basic mathematics activities (around 10 hours per week) in 3 months between March and June 2018 supported by 12,735 facilitators, 86 percent of them were teachers (ibid). Once these remedial activities were completed, correct responses on student basic math assessment significantly improved from a starting point (baseline) of around 40 percent to 67 percent on average (end-line) (ibid). As evidence for the impact of scaling-up, these figures are highly inadequate. First, this is a before-after comparison, not controlling for changes over time that were not related with the project. Furthermore, the data quality is not satisfactory[10]. Therefore, we estimate the impact in a more rigorous way, as explained in the

---

8 The SMC federation is a network of SMCs in a geographical or administrative area (municipality or district) that supports SMCs in its network.

9 Each SMC federation organized a general assembly to report and discuss the result of forum to member SMCs, and to pledge to achieve the target. Subsequently, each SMC organized a community general assembly to discuss the objective, planning and implementation arrangement of remedial activities.

10 SMC conducted basic math assessment and reported the result to local education office, which complied and forwarded them to the Ministry of Primary Education. There are several potential biases in

next section.

3. Evaluation strategy

(1) Sampling of schools and student test data

Of the 3,655 public primary schools in Tillaberi region, 3,519 schools in 13 departments taught at least one class of 1st to 4th grades. Niger has two types of public primary school: those that teach in French (plus local language); and (ii) those that teach Arabic in addition to the other subjects. The latter type is called “Franco-Arab primary school”, which constitute around 10 percent of public primary schools in the region.

The EPT project selected 31 public primary schools which teaches subjects in French (and local language) and 3 Franco-Arab primary schools. Among the former, 28 schools were selected through three-stage random sampling[11]: at the departmental, commune, and school levels. The other remaining three were schools where the EPT project piloted upgraded version of PMAQ which integrates basic reading and writing in addition to basic math component[12]. The three Franco-Arab primary schools were randomly chosen from the same departments selected in the process of above-mentioned random sampling. As a result, the 34 schools, consisting of 31 public primary schools and 3 Franco-Arab primary schools, are our evaluation sample.

[Insert Table 1-1: Distribution of sampled schools by department and commune]

the student assessment results reported by each SMC, and compiled by the Ministry of Primary Education. First, as the math test was voluntarily organized by each SMC, the quality of controlling test organization varies among SMCs. Secondly, since it was not mandatory for SMCs to submit activity report to the local education administration office, all the reports were not collected. Third, the numbers in SMC activity report were summed up at the local education administration offices. In the process, the number might not be correctly calculated, because of large volume of assessment results. In addition to those potential biases, the student level individual test result is not available in the data compiled by the Ministry. Given those limits, this paper uses the data collected by local NGO under the contract with the EPT project.

11 Five departments situated mainly in northern part of the region were excluded from sampling frame because of security concerns.

12 For the three schools, math workbooks were delivered by the project. On the other hand, other schools came to local education office to receive workbooks. The transportation expenses were born by the multilateral support.

Local NGO conducted three rounds of survey: in March 2017 (pre-baseline); December 2017 (baseline); and June 2018 (end-line)[13]. As explained in the previous section, due to the delay of math workbooks procurement, the schedule of scaling up PMAQ was postponed. Then, the first survey on March 2017 which was initially conducted as baseline survey became pre-baseline survey. As the school calendar in Niger starts from October and finishes in June next year, the pre-baseline survey was conducted in the 2016-17 school year, and baseline and end-line surveys were organized in the 2017-18 school year.

At each round of surveys, teams of local NGO organized basic math tests to measure learning outcomes of all grades students at each school[14]. The test was composed for each grade as follows:

- Grade 1: (i) single-digit addition (12 questions);
- Grade 2: (i) plus (ii) single-digit subtraction (12 questions);
- Grade 3: (i) and (ii) plus (iii) two-digit addition including problems with carrying and subtraction including problems with borrowing (11 questions) plus (iv) single-digit multiplication (9 questions);
- Grade 4: (i), through (iv) plus (v) three-to-four-digit addition including problems with carrying and subtraction including problems with borrowing (12 questions); and
- Grades 5 and 6: (i) through (v) plus (vi) single-digit division including problems with remainders (5 questions).

There were 44 test items for 3rd grade students, 56 for 4th grade students, and 61 for 5th and 6th grade students. We used 56 items for 4th grade students to compare with 4th and 5th grade students' test scores.

---

[13] The authors were not involved in the survey design and data collection process. School sampling, preparation of test items and surveys were conducted by the EPT project and local NGO. JICA provided all data used in our study.

[14] The test took 30 minutes for 1st and 2nd grades, and 45 minutes for the other grades.

(2) Constructing panel data

Since the three rounds of survey were not designed to track same students, we constructed panel data of 4th through 6th grade students ex post by manually linking data in the three rounds of survey. As student name was typed at each round of survey, it caused inconsistency of spelling among the three-rounds. For overcoming such inconsistency of spelling in student name, we utilized fuzzy-matching method (Robinson, 2019). Specifically, we took four steps to construct panel data from the three rounds of survey data. First, we removed accent marks above letters in student names, and transformed student and school name to small letters.

Next, we assigned identification numbers to each student in the end-line data. Personal names in Niger generally consist of an individual first name, father's first name and grandfather's first name. Some students were registered by all of them, but the others were only by own first name and father's name. Thus, we fuzzy-matched end-line data with baseline data using letters of the student first name, father's name and school name by grade[15]. If there were students with exactly same name, we used grandfather's name and matched it again. In case that students still had same name, we excluded it[16]. As a result, the panel data of two data points (baseline and end-line) include 1,068 students from 4th through 6th grade.

Third, we fuzzy-matched our panel data (baseline and end-line) with pre-baseline data for the corresponding previous grade or the same grade during the 2016-17 school year. For example, we fuzzy-matched 4th grade student in the 2017-18 school year with 3rd grade in the previous school year. Considering that students might have repeated the grade, we fuzzy-matched the remaining unmatched data with data of 4th grade student in the 2016-17 school year. Our panel data of three data points contains 710 students.

None of the three rounds of survey collected student gender information. We recovered the student gender information based on the typical correspondence of student first name and gender. We observed around 800 types of student first name in total in the pooled data

15 The degree of fuzziness allowed one-letter differences in first name and father's name.

16 Five pairs of students had same first and family name and belong to the same school. We excluded them from the panel data sample.

from baseline and end-line surveys[17].

The trends in average math test scores of 4th and 5th grade student group in three rounds of survey is shown in Graph 1. “4th grade student group” is those who were 4th grade students in the 2017-18 school year, which includes students who moved up from 3rd grade or repeated 4th grade. “5th grade student group” is those who were 5th grade in 2017-18, which includes students who moved up from 4th grade or repeated 5th grade. The trends in average math test scores of 4th and 5th grade student groups are parallel from pre-baseline to baseline survey. On the other hand, the 4th grade average math test scores improved more than 5th grade student group from baseline to end-line survey, which we will analyze in the following section. Table 1-2 presents descriptive statistics of the test results.

[Insert Graph 1: Trends in average math test scores (4th to 6th grade groups),

and Table 1-2: Number of students and test scores by grade group]

(3) Estimation Strategy

As noted, because of budget constraint, the distribution of math workbooks was limited to students from 1st to 4th grades. That discontinuity of intervention between 4th and 5th grades allows us to estimate the impact of PMAQ by equation (1) using the pooled data from baseline and end-line surveys.

$$Y_{ijt} = \alpha + \beta_1\ Grade4_i + \beta_2\ Post_t + \lambda\ DID_{it} + X_i\gamma_1 + X_j\gamma_2 + \varepsilon_{ijt}\,, \quad (1)$$

Y_{ijt} is the raw test score or its standardized value of student i in school j at the timing of tth survey; $Grade4_i$ is a dummy variable that takes 1 for the student i of 4th grade in the 2017-18 school year, and takes 0 for the 5th grade; $Post_t$ is a dummy variable that takes 1 for the end-line

17 The EPT project provided association between student first name and gender. Boys’ and girls’ names are rarely gender-ambiguous in Niger.

data; DID_{it} is an interaction term of $Grade4_i$ (treatment) and $Post_t$, which identifies the impact of PMAQ. X_i is a vector of student characteristics for student i (e.g., the female dummy); X_j is a vector of school characteristics for school j (e.g., the dummy for Franco-Arab school or pilot school of basic reading) and writing activity; ε_{ijt} is a residual of the regression. Scalers of parameter α, β_1, β_2 and λ, and vectors of parameters γ_1 and γ_2 are to be estimated econometrically. The standard errors are clustered at school level.

As the main interest of this study is not on the impact of school characteristics on learning, we replaced $X_j\gamma_2$ in the equation (1) with school fixed effects, α_j. The equation (1) becomes

$$Y_{ijt} = \alpha_j + \beta_1 \, Grade4_i + \beta_2 \, Post_t + \lambda \, DID_{it} + X_i\gamma_1 + \varepsilon_{ijt} , \quad \text{(1-a)}.$$

By using the balanced panel data of two periods instead of the unbalanced pooled data, we controlled the student characteristics more flexibly, using student fixed effects, α_i. Note that each student attends the same school in both baseline and end-line. In other words, student fixed effects and school fixed effects are perfectly collinear, rendering the latter redundant. The estimation model becomes

$$Y_{ijt} = \alpha_i + \beta_2 \, Post_t + \lambda \, DID_{it} + \varepsilon_{ijt} , \quad \text{(2)}.$$

By construction, parameters β_1, γ_1, and γ_2 in the equation (1) cannot be identified.

The assumption under which we identify λ is parallel trends in the test scores of 4th and 5th grade student groups before the scaled-up of PMAQ. If the parallel trend assumption does not hold, λ cannot correctly identify the impact of PMAQ. The parallel trend assumption is checked by the equation (3) using the pre-baseline and baseline data. $Pre\text{-}Treatment_t$ is a dummy variable which takes 1 for the baseline data; DID'_{it} is an interaction term of $Grade4_i$ and $Pre\text{-}Treatment_t$. If the parallel trend assumption holds, λ' will be close to zero and not be statistically significant.

$$Y_{ijt} = \alpha_i + \beta_2\ Pre\text{-}Treatment_t + \lambda'\ DID'_{it} + \varepsilon_{ijt}\ , \quad (3)$$

As another test for our parallel trend assumption, we also check the trend of basic math learning of students in different grades, but not using math workbooks in remedial activity. More concretely, we estimate equations (4) and (5) comparing 5th and 6th grade students in 2017-18. In estimating (4), we use the pooled data of 5th and 6th grade students, while in estimating (5), we use the panel data of 5th and 6th grade students. $Grade5_i$ is a dummy variable which takes 1 for 5th grade students in 2017-18, and takes 0 for the 6th grade; DID''_{it} is an interaction term of $Grade5_i$ and $Post_t$. If the trends in test scores of those grade students are parallel, λ'' will be close to zero and not be statistically significant.

$$Y_{ijt} = \alpha_j + \beta_1\ Grade5_i + \beta_2\ Post_t + \lambda''\ DID''_{it} + X_i\gamma_1 + \varepsilon_{ijt}\ , \quad (4)$$

$$Y_{ijt} = \alpha_i + \beta_2\ Post_t + \lambda''\ DID''_{it} + \varepsilon_{ijt}\ , \quad (5).$$

4. Results

(1) Average impact on basic math learning

The average impact of PMAQ on student basic math learning is estimated at 5.6 score points out of the total of 56 points using the pooled data and equation (1) (Table 2-1 (I)), 5.9 score points using equation (1-a) which controls school fixed effect (Table 2-1 (II)), and 6.0 score points when panel data is employed (Table 2-1 (III) and (IV)). All the estimated values are statistically significant at the 1 percent level. Then, we standardized the impact with the mean and the standard deviation of 4th grade students baseline score. The impact is 0.38 standard deviations using the pooled data (Table 2-2 (II)) and the panel data (Table 2-2 (III) and (IV)).

[Insert Table 2-1 and 2-2: Average impact of PMAQ on math learning]

Spider charts of Graph 2-1 to 2-3 display the before/after comparison of correct response rates for each test item among 4th grade students. Correct response rates of each item improved from baseline to end-line, but the rates are still low in items No. 31 and 32 (two-digit subtraction with borrowing), and items from No. 52 through 56 (three to four-digit addition with carrying and subtraction with borrowing). The test items that exhibits improved correct response rate differ according to the baseline score level. For example, students with baseline score less than the median correctly answered more single addition and subtraction items at the end-line survey (Items No. 1 through 12 on Graph 2-2). Students with baseline score over the median correctly answered more test items involving single-digit multiplication (Items No. 36 through 44) and three-digit addition and subtraction without borrowing (Items No. 45 through 51 on Graph 2-3).

[Insert Graph 2-1 to 2-3: Correct response rate by math item (4th grade students)]

(2) Testing parallel trends

The trends in average scores of 4th and 5th grades student group from pre-baseline to baseline survey are nearly parallel (Graph 1). The parallel trends are checked by the equation (3) using pre-baseline and baseline scores in the panel data. The parameter estimate for λ' is -0.92 score points out of the total of 44 points which is not statistically significant (Table 3 (I))[18]. It shows that the parallel trend assumption holds for identifying the impact of PMAQ through the difference in differences (DID) of 4th and 5th grades student test scores.

[Insert Table 3: Regression results of parallel trend test (raw test score)]

Next, we checked the trend of basic math learning of students who did not use math workbooks

[18] Most 4th grade students were at 3rd grade at the timing of pre-baseline survey. As explained in 4. (2), the maximum scale for 3rd grade students is 44.

in remedial activity by comparing 5th and 6th grade student test scores. If the trends of test score of 5th and 6th grade students are different, the DID of 4th and 5th grade student test scores does not accurately identify the impact of PMAQ. The results from equation (4) and (5) are shown in Table 3 (II) and (III). The parameter estimate for λ'' is 1.00 score point using the pooled data and 0.47 score point using the panel data. Neither are statistically significant. Those results indicate that the DID of the test scores of 4th and 5th grades identifies the impact of PMAQ on student math learning.

(3) Robustness check of estimated impact

To check the robustness of the estimated impact of PMAQ further, we added interaction terms of X_i and $Post_t$, and those of X_j and $Post_t$ in equations (1) and (1-a) as below,

$$Y_{ijt} = \alpha + \beta_1\ Grade4_i + \beta_2\ Post_t + \lambda\ DID_{it} + X_i\gamma_1 + X_j\gamma_2 + (X_i \times Post_t)\gamma_3 + (X_j \times Post_t)\gamma_4 + \varepsilon_{ijt}, \quad \text{(1-b)}$$

$$Y_{ijt} = \alpha_j + \beta_1\ Grade4_i + \beta_2\ Post_t + \lambda\ DID_{it} + X_i\gamma_1 + (X_i \times Post_t)\gamma_3 + \varepsilon_{ijt}, \quad \text{(1-c).}$$

The estimated values of λ from equations (1-b) and (1-c) remain almost at the same level as reported in the previous sub-section (Table 4-1 (I) to (III)).

[Insert Table 4-1: Regression result of robustness check (1) (raw test score)]

Given the structure of the basic math test, students with higher baseline scores would find it harder to gain scores at the end-line survey than students with lower baseline scores. Since average baseline scores of 4th grade students are below those of 5th grade students, the difference of the baseline score level might engender a bias in estimating the impact of PMAQ. In order to check such potential bias, we conducted a regression with the sub-samples in the same range of baseline test scores from 10 to 37 score points, equivalent to 2nd and 3rd quartile of the 4th grade baseline scores. The sub-sample of 5th grade students in the bandwidth represents around half of the panel data sample[19]. The test scores are standardized with the mean and standard deviation of the whole sample of the 4th grade student data. The estimated values of λ with the sub-sample is 0.28 to 0.29, which is slightly smaller than the values with the whole sample (Table 4-2) but still statistically significant.

[Insert Graph 3-1: Cumulative density curve of test scores of 4th grade student group,
Graph 3-2: Cumulative density curve of test scores of 5th grade student group, and
Table 4-2: Regression results of robustness check (2)]

Although the distribution of math workbooks was limited to 1st to 4th grade students, some schools organized supplementary classes for 5th and 6th grade students. The supplementary classes for 5th and 6th grade students might have targeted basic math without using math workbooks. To check the robustness of the estimated impact of PMAQ, we conducted a regression with 4th grade students, and the sub-samples of 5th grade students of schools where supplementary classes for the grade were organized. The test scores are standardized with the mean and standard deviation of the whole sample of the 4th grade student data. The estimated value of λ with the limited sample is 0.32, which is slightly below the values with the whole sample (Table 4-3) but still statistically significant.

[Insert Table 4-3: Regression results of robustness check (3)]

(4) Heterogeneous impacts

We analyzed the heterogeneity of impacts of PMAQ on student math learning with respect to baseline score level and gender. If the impact is larger on students with lower baseline test

[19] The percentage of 5th grade students with baseline scores below 9 score points is 9 percent of the original 2 data point panel data sample. The percentage of 5th grade students with baseline below 37 score points is 57 percent of the original two-data-point panel data sample.

scores, it indicates that PMAQ helps those students to catch up their basic math understanding and skills. If the impact is larger on girls than on boys, it shows that PMAQ reduces gender disparity in basic math learning. As shown in Graph 4, the average baseline score of girls is lower than boys.

[Insert Graph 4: Boxplot of baseline test score by sex and grade (4th and 5th grade students), and Table 5: Number of 4th and 5th grade students by test score level and sex]

Because of the limited number of samples, we expanded the equation (1-a) for the heterogeneity analysis as

$Y_{ijt} = \alpha_j + \beta_1\ Grade4_i + \beta_2\ Post_t + (\lambda_0 + \lambda_1\ Female_i)\ DID_{it} + X_i\gamma_1 + \varepsilon_{ijt}$, (6-a), and

$Y_{ijt} = \alpha_j + \beta_1\ Grade4_i + \beta_2\ Post_t + (\lambda_0 + \lambda_{1L}\ Lowest\_Score_i + \lambda_{1M}\ Median\_Score_i) + DID_{it} + X_i\gamma_1 + \varepsilon_{ijt}$, (6-b).

By using balanced panel data, the heterogeneous impact will be accurately estimated. In the equation (6-a), parameter λ_1 will identify the treatment effect additional for girls in comparison with boys. In the equation (6-b), sample students are split into three groups: lowest baseline score group (the 1st quartile of the score of each grade); median baseline score group (the 2nd and 3rd quartiles of the score of each grade); and highest baseline score group (4th quartile of the score of each grade), and then we define $Lowest\_Score_i$ as the dummy variable which takes 1 if the test score of student i falls into the lowest baseline score group and $Median\_Score_i$ as the dummy variable which takes 1 if the baseline test score of student i falls into the median baseline score group. Parameter λ_{1L} andλ_{1M} will identify the treatment effect additional for students in different groups in comparison with the highest baseline score group.

As the female dummy and lower baseline score dummy are correlated, we included both in a same equation. The heterogeneity impact model (6) can be expanded as

$Y_{ijt} = \alpha_j + \beta_1\ Grade4_i + \beta_2\ Post_t + (\lambda_0 + \lambda_{1L}\ Lowest\_Score_i + \lambda_{1M}\ Median\_Score_i +$

$\lambda_2\ Female_i) DID_{it} + X_i\gamma_1 + \varepsilon_{ijt}$, (7), and

$Y_{ijt} = \alpha_j + \beta_1\ Grade4_i + \beta_2\ Post_t + (\lambda_0 + \lambda_{1L}\ Lowest\_Score_i + \lambda_{1M}\ Median\_Score_i +$

$\lambda_2\ Female_i + \lambda_{3L}\ Lowest\_Score_i \times Female_i + \lambda_{3M}\ Median\_Score_i \times Female_i)\ DID_{it} +$

$X_i\gamma_1 + \varepsilon_{ijt}$, (8)

from which we can test whether there is an interaction effect of two types of disadvantages combined through examining parameter λ_3.

The impact of PMAQ is larger for lower baseline score groups (Table 6-1 (II), (IV) and (V)). The estimated value of λ_{1L} is 0.58 standard deviations, and that of λ_{1M} is 0.21 standard deviations, even after controlling interaction term of the DID and gender (Table 6-1 (V)). PMAQ helped children with lower baseline scores catch up basic math learning with higher baseline score group. On the other hand, the impact additional for girls is not statistically significant (Table 6-1 (IV) and (V)).

[Insert Table 6-1: Heterogeneous impacts of PMAQ on math learning (standardized test score)]

To examine the robustness of heterogeneous impacts, we added the interaction terms of X_i and $Post_t$ to equations (6) to (8). For instance, when an interaction term of the female dummy and $Post_t$ is added, the DID heterogeneity term of girls shows the impact of PMAQ additional on 4th grade girls in comparison to 5th grade girls, independent of the additional impact on boys. Furthermore, as already discussed, the estimated value of λ might be biased if trends in scores without interventions are different depending on the baseline score level. For these reasons, we replaced $Lowest\_Score_i$ with $Lowest\_Score'_i$ (taking 1 when the test score of student i is 0 to 9 score points, which is the range for 1st quartile baseline score of 4th grade students), and also

$Median\_Score_i$ with $Median\_Score'_i$ (taking 1 when the test score of student i is 10 to 37 score points, which is the range of 2nd and 3rd quartile of 4th grade students) in equations (6) to (8). We also added interactions of $Post_t$ and the female dummy, $Lowest\_Score'_i$ and $Median\_Score'_I$ to the equations.

The estimation results are reported in Table 6-2. The impact on the highest score group (λ_0) is positive and statistically significant at the 1 percent level (Table 6-2 (II)). 4th grade students of the highest score group learned basic math better than 5th grade students of the same bandwidth of the score. On the other hand, the impact on the lowest score group (λ_{1L}) is positive but not statistically significant (Table 6-2 (II) and (V)). The lowest score group of 4th grade as well as 5th grade students caught up with the highest score group of each grade.

The average baseline score of girls is lower than boys, and the time trend in 5th grade girls' learning is negative (Table 6-2 (III), (IV) and (V)), which indicates that the gender disparity in learning outcomes tends to become larger. The estimated value of λ_2 is positive and statistically significant at the 5 percent level (Table 6-2 (III) to (V)), which indicates that 4th grade girls could learn basic math better than 5th grade girls by using math workbooks in the remedial activity.

[Insert Table 6-2: Robustness check of heterogeneous impacts of PMAQ on math learning (standardized test score)]

5. Discussions

PMAQ improved basic math learning even in a context of limited capacity of educational administration and teachers. PMAQ helped students with lower baseline score catch up basic math learning. The remedial activity is organized not by grade or age, but by student proficiency level. Students could learn math with workbooks which matches with their proficiency levels, supported by activity facilitators. The learning contents are standardized by math workbooks, which enabled scaling up remedial activity with quality in a difficult context. The remedial activity using math workbooks also helped teachers to reflect their teaching practices in daily classes (Hara et al., forthcoming). Teachers tended to have a negative perception on student learning ability, since most of them could not follow daily classes. In remedial activities of PMAQ, teachers checked student answers in math workbooks and gave them an instruction individually. Throughout the process, teachers noticed the progress of student math learning, and changed positively their perception on student learning ability. Students also felt teachers closer to them, and asked more questions (ibid).

The EPT project has continuously developed a package of interventions for addressing various educational needs in local communities in Africa. Kozuka et al. (2016) found that a package of trainings for school principals and SMC representatives increased student enrollment, decreased student repetition, and lowered teacher absence through establishing community-wide collaboration in Burkina Faso. The intervention also enhanced social capital in the community (Sawada et al., 2016). A combination of grant provision and training for school principals and SMC representatives further improved student learning outcomes in French and math in Niger (Kozuka, 2018). This study contributes to the literature on the empirical study on the EPT approach in Africa through providing rigorous evaluation of scaled up intervention of the novel package of interventions, PMAQ.

While PMAQ helped students with lower baseline scores improve basic math learning, it was not sufficient to decrease gender disparity in learning. Girls in Niger socially mature earlier than boys, and hesitate to ask questions or pose answer in daily class, because they worry what the others think of them. Such behavior hampers girls' math learning, because they receive less feed-back from teachers. In the extra-curricular remedial activity of PMAQ, where activity facilitators check math workbooks individually, girls can learn basic math by their own paces and ask questions. This study investigated the heterogeneity of impacts with respect to gender. We found that 4th grade girls who learned with the workbook improved math learning than 5th grade girls who did not; however, the gender disparity in basic math learning still remains. In Niger, the percentage of girls who reach to the last grade of primary education in the age group is 65 percent, which is lower than boys by 13 percentage points in 2016 (UIS, 2020). Low learning achievement is a major reason of dropout of girls from primary education (Perlman et al., 2016). PMAQ has room to improve for reducing gender disparity in learning.

6. Conclusion

The scaling-up of PMAQ in Niger's Tillaberi region tried to help students improve basic math learning through extra-curricular remedial activity. PMAQ aims to improve basic math learning by two types of interventions: a one-day training of SMC president and secretary to strengthen the capacity of SMC to organize basic math remedial activity, and a set of distribution of math workbooks and a two-day training of activity facilitators. Because of budget constraint, the distribution of math workbooks was limited to 1st to 4th grades. Focusing on the discontinuity of intervention between 4th and 5th grades, we investigated the impact of PMAQ on student math learning using three rounds survey data. The average impact is estimated to be 0.36 to 0.38 standard deviations. The impact is larger for students with lower baseline score, which indicates that the remedial activity helped students left behind catch up basic math learning.

In developing countries, a mismatch between curricula and the learning levels of students expands as curricular progresses. While most students are left behind, teachers are required to finish prescribed syllabus. As a result, most of children finish or drop out from primary education without mastering foundational skills for learning: basic reading and math. The scaling up of PMAQ in Niger shows that once children have a chance of learning which matches with their proficiency levels, they can improve math learning. Governments in sub-Saharan Africa should pursue measures that help children master foundational skills for overcoming learning crisis.

Reference

Abadzi, Helen, Stavri Llambiri. (2011). "Selective Teacher Attention in Lower-Income Countries: A Phenomenon Linked to Dropout and Illiteracy?" *Prospects*, 41, 491–506.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobini Mukerji, Marc Scotland, and Michael 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.

Banerji, Rukmini. (2000). "Poverty and Primary Schooling: Field Studies from Mumbai and Delhi." *Economic and Political Weekly*, 35 (10), 795-802.

Bashir, Sajitha, Marlaine Lockheed, Elizabeth Ninan, and Jee-Peng Tan. (2018). Facing Forward: Schooling for Learning in Africa. Africa Development Forum Series. Washington DC: World Bank.

Bold, Tessa, Deon Filmer, Gayle Martin, Ezequiel Molina, Brian Stacy, Christophe Rockmore, Jakob Svensson and Waly Wane. (2017). "Enrollment without Learning: Teacher Effort, Knowledge and Skill in Primary Schools in Africa." *Journal of Economic Perspectives*, 31 (4): 185-204.

Evans, David K. and Anna Popova. (2016). "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *World Bank Research Observer*, 31 (2), 242-270.

Hara, Masahiro, Takao Maruyama, Akiko Kageyama and Nobuhiro Kunieda. (Forthcoming). Quality Learning through Community-wide Collaboration: A Methodology to Overcome the "Learning Crisis" in Niger. In M. Nishimura (Ed.), *Community participation with schools in developing countries: Towards equitable and inclusive basic education for all*. Routledge Research in Educational Equality and Diversity Series. New York & Oxon: Routledge.

JICA. (2017). Final Evaluation Report on School for All: The project on support to educational development through community participation. Tokyo: Japan International Cooperation Agency.

Kozuka, Eiji, Yasuyuki Sawada and Yasuyuki Todo. (2016). “How Can Community Participation Improve Educational Outcomes? Experimental Evidence from a School-Based Management Project in Burkina Faso.” *JICA-RI Working Paper* No.112. Tokyo: JICA Research Institute.

Kozuka, Eiji. (2018). “Enlightening Communities and Parents for Improving Student Learning Evidence from Randomized Experiment in Niger.” *JICA RI Working Paper.* No.166. Tokyo: JICA Research Institute.

Kunieda, Nobuhiro, Takao Maruyama, Akiko Kageyama and Masahiro Hara. (Forthcoming). Educational Development through Community-wide Collaboration: How to Establish a Sustainable Community-wide Initiative to Improve Education. Nishimura. *op.cit*.

Ministry of Primary Education in Niger. (2016). *STATISTIQUES DE L’ÉDUCATION DE BASE ET ALPHABÉTISATION ANNUAIRE: ANNUAIRE 2015-2016*. Niamey: The Ministry of Primary Education in Niger.

Ministry of Primary Education in Niger. (2018). [Situation de la mise en œuvre du PMAQ dans les écoles (BILAN)]. Unpublished Raw Data.

Ministry of Primary Education in Niger. (2019). *STATISTIQUES DE L’ÉDUCATION DE BASE ET ALPHABÉTISATION ANNUAIRE*. Niamey: The Ministry of Primary Education in Niger.

Organisation Nigérienne des Educateurs Novateurs (ONEN). (2017). *Rapport final de l’étude à mi-parcours dans 45 écoles primaires de la région Tillaberi dans le cadre de l’evaluation des résultats des interventions du Projet Ecole Pour Tous Phase II (Mars-Avril 2017)*. Niamey: JICA.

Organisation Nigérienne des Educateurs Novateurs (ONEN). (2018a). *Rapport final de l’étude à mi-parcours dans 45 écoles primaires de la région Tillaberi dans le cadre de* *l'evaluation des résultats des interventions du Projet Ecole Pour Tous Phase II (Décembre 2017-Janvier 2018)*. Niamey: JICA.

Organisation Nigérienne des Educateurs Novateurs (ONEN). (2018b). *Rapport final de l'étude à mi-parcours dans 45 écoles primaires de la région Tillaberi dans le cadre de l'evaluation des résultats des interventions du Projet Ecole Pour Tous Phase II (Juin-Juillet 2018)*. Niamey: JICA.

PASEC. (2016). *PASEC 2014 – Performances du système éducatif nigérien: Compétences et facteurs de réussite au primaire*. PASEC. CONFEMEN. Dakar.

Perlman, Daniel, Fatima Adamu, and Quentin Wood. (2018). "Why Do Adolescent Girls Drop Out of School in Niger? A Combined Quantitative and Qualitative Analysis." *Marché et organisations*, 2 (32), 179-194.

Pritchett, Lant, Amanda Beatty. (2015). "Slow Down, You Are Going Too Fast: Matching Curricula to Student Skill level." *International Journal of Educational Development*, 40, 276-288.

Robinson, David (2019). *fuzzyjoin: Join Tables Together on Inexact Matching. R package version 0.1.5*. https://CRAN.R-project.org/package=fuzzyjoin.

Sawada, Yasuyuki, Takeshi Aida, Andrew S. Griffen, Eiji Kozuka, Haruko Noguchi, and Yasuyuki Todo. (2016). "Election, Implementation, and Social Capital in School Based Management: Evidence from a Randomized Field Experiment on the COGES Project in Burkina Faso". *JICA-RI Working Paper No.120*. Tokyo: JICA Research Institute.

UIS (UNESCO Institute for Statistics). (2020). Data for the Sustainable Development Goals: Niger. Retrieved from http://uis.unesco.org/country/NE (Access date: April 23, 2020)

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

World Bank. (2017). *Republic of Niger: Niger service delivery indicators education 2015*. Washington DC: World Bank.

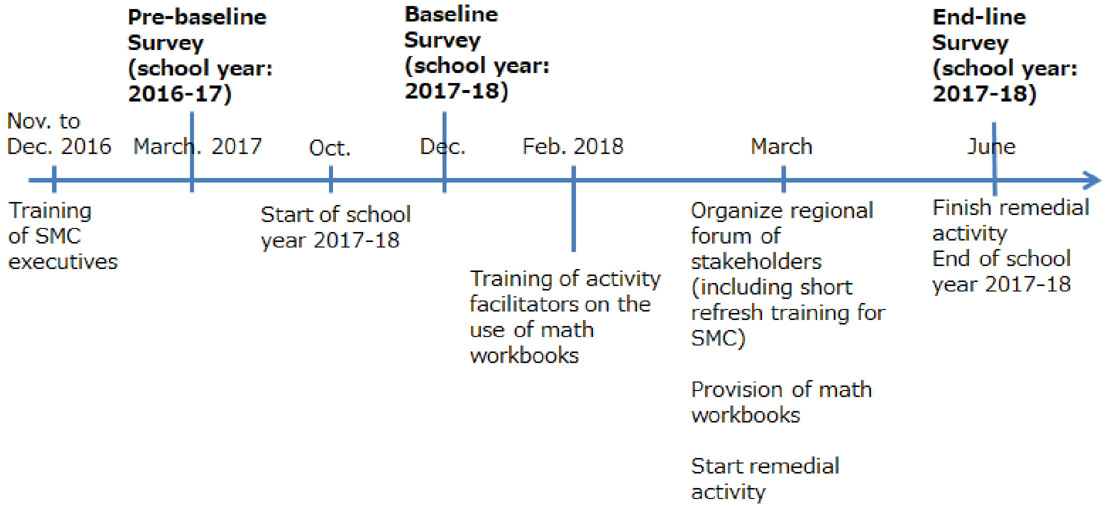
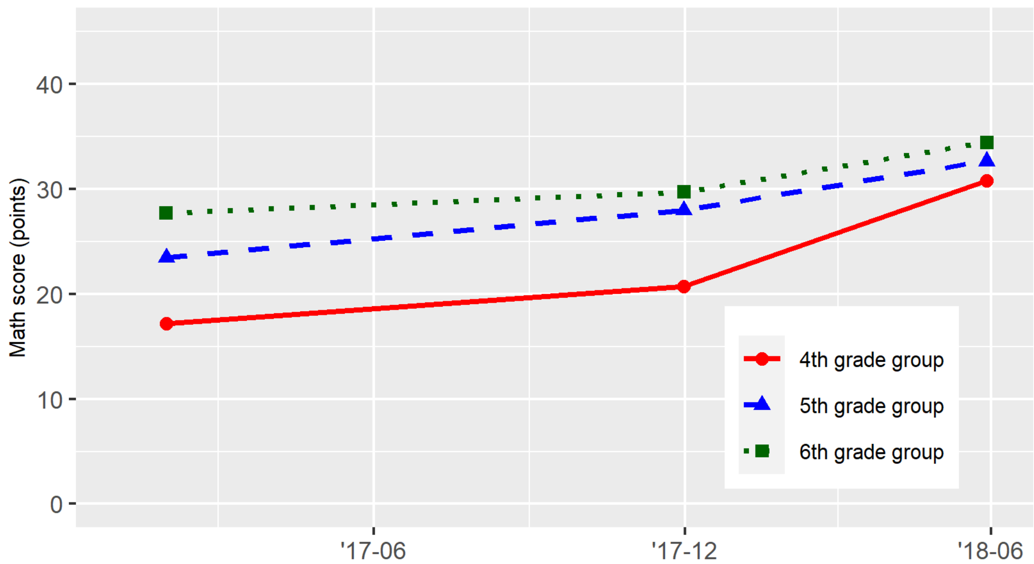
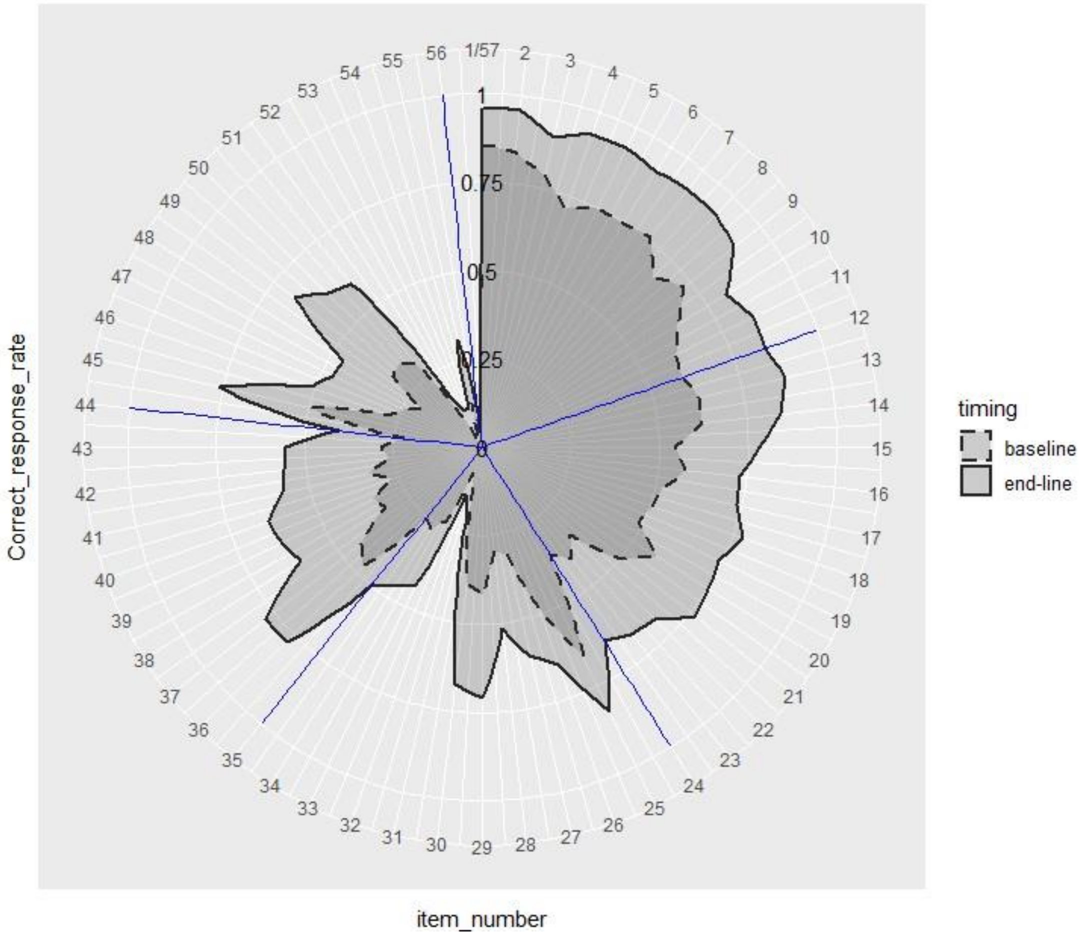


Figure 1: Timeline of scaled-up intervention

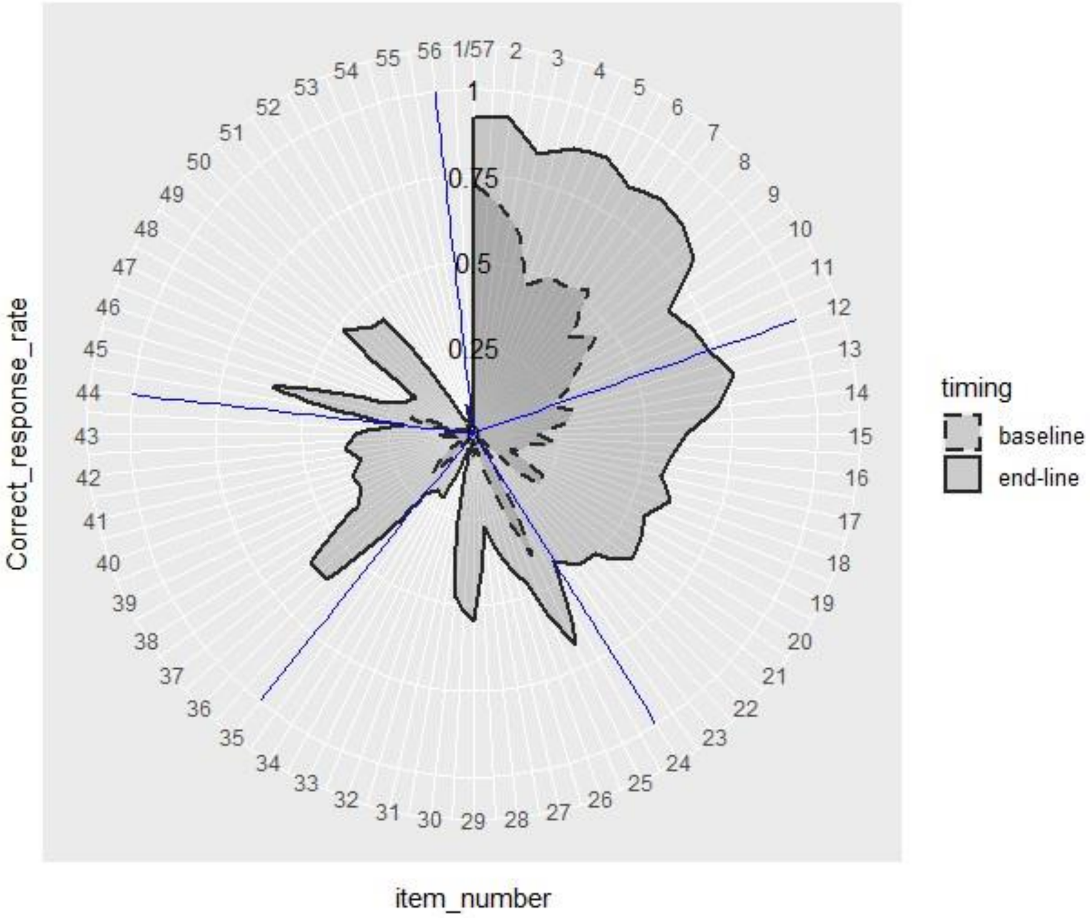


N.of 4th grade group: 240, N. of 5th grade group: 220,
N. of 6th grade group: 242. Maximum Scale of test score is 44.

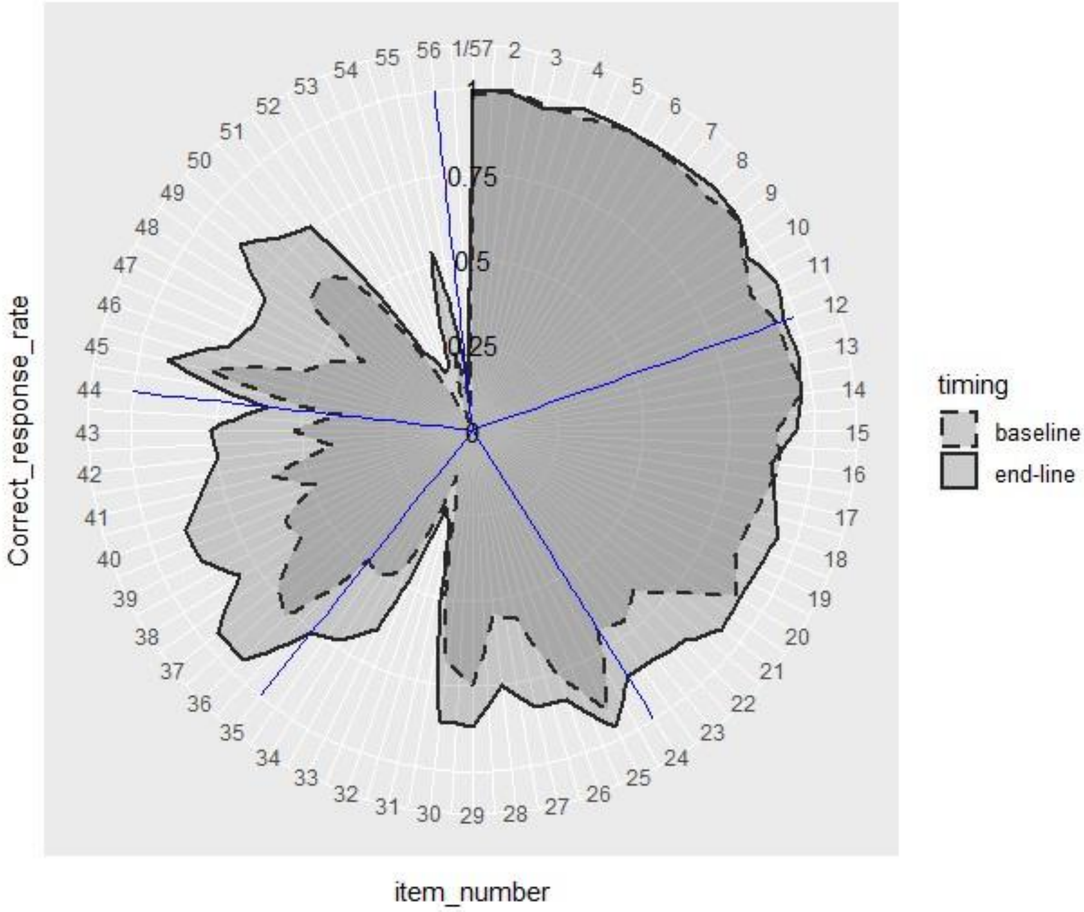
Graph 1-1: Trends in average math test scores (4th to 6th grade groups)



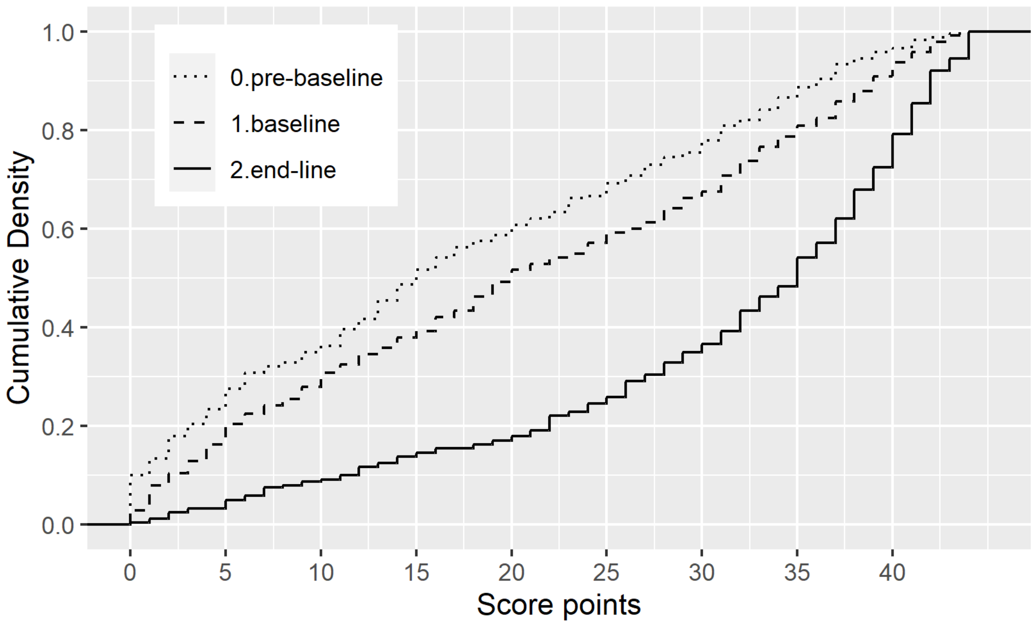
Graph 2-1: Correct response rates by item (4th grade students)



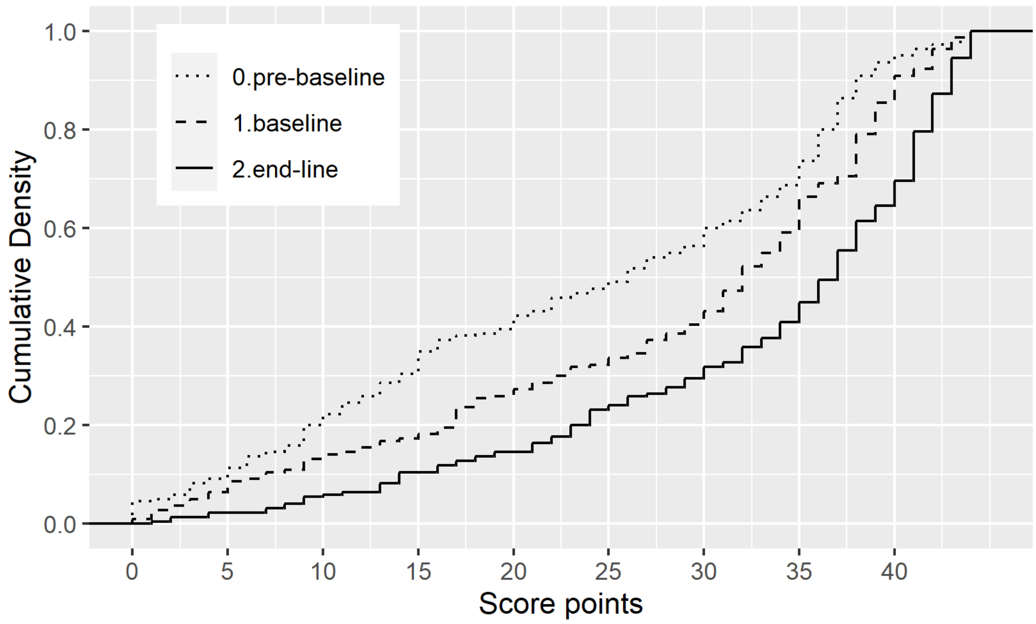
Graph 2-2: Correct response rates by item
(4th grade students below median baseline score)



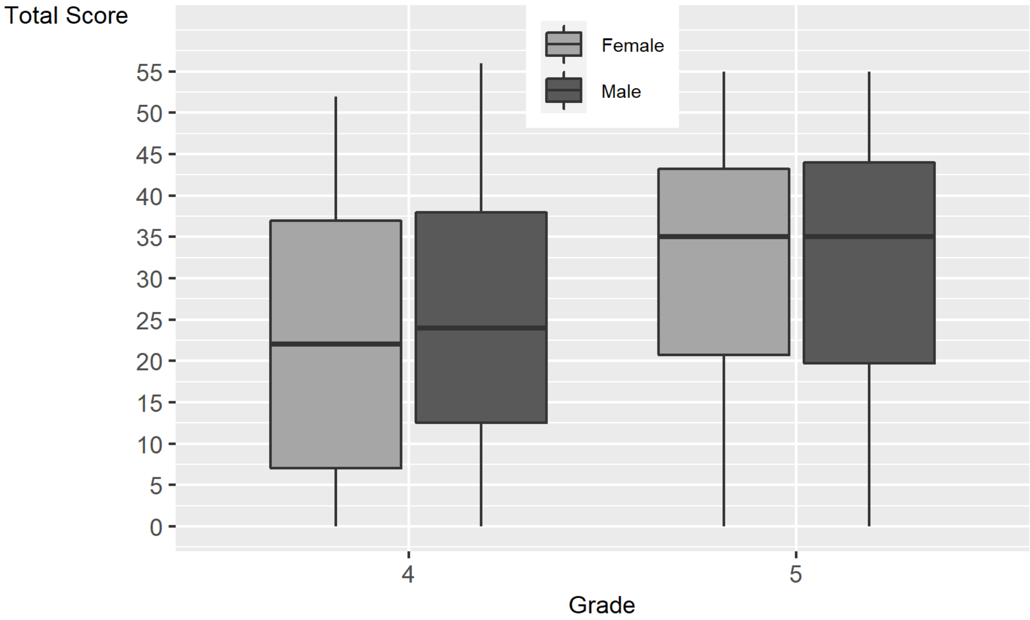
Graph 2-3: Correct response rates by item
(4th grade students over median baseline score)



Graph 3-1: Cumulative density curve of test score of 4th grade student group



Graph 3-2: Cumulative density curve of test score of 5th grade student group



Graph 4: Boxplot of baseline test score by sex and grade (4th and 5th grade group)

Types of public primary school	Department	Commune	N. of schools which have both 4th & 5th grades (t1 and t2) (A)	N. of schools which have only 4th grade (t1 and t2) (B)	N. of schools which have only 5th grade (t1 and t2) (C)	N. of schools of evaluation sample of this study (t1 and t2) (A)+(B)+(C)
Public primary school which teaches subjects in French or local language	Kollo	Kollo	2	4	0	6
		Hamdalaye	2	1	1	4
	Say	Say	3	1	0	4
		Gueladio	0	1	3	4
	Gotheye	Gotheye	1	2	0	3
		Dargol	2	1	0	3
	Torodi	Makalondi	2	1	1	4
		Torodi	3	0	0	3
Franco-Arab	Kollo	Kollo	1	0	0	1
		N'Dounga	0	0	1	1
	Torodi	Torodi	0	1	0	1
Total			16	12	6	34

Note t1: baseline survey; t2: end-line survey

Table 1-1: Distribution of sampled schools by department and commune

Pooled Data (2t): Maximum scale (full mark) of test score is 56

	Pre-baseline		Baseline		End-line	
	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)
Grade 4 group			545 (288)	22.0 (15.4)	516 (290)	34.3 (14.7)
Grade 5 group			521 (237)	31.0 (14.7)	413 (199)	37.7 (14.2)
Grade 6 group			571 (276)	35.1 (14.6)	536 (272)	40.6 (14.1)

Panel Data (2t): Maximum scale (full mark) of test score is 56

	Pre-baseline		Baseline		End-line	
	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)
Grade 4 group			360 (197)	23.7 (15.7)	360 (197)	36.0 (14.2)
Grade 5 group			332 (152)	32.0 (14.5)	332 (152)	38.3 (14.0)
Grade 6 group			376 (196)	35.0 (14.4)	376 (196)	40.8 (14.0)

Panel Data (3t): Maximum scale (full mark) of test score is 56

	Pre-baseline		Baseline		End-line	
	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)
Grade 4 group			240 (125)	23.1 (16.1)	240 (125)	35.7 (14.6)
Grade 5 group			220 (102)	32.1 (15.2)	220 (102)	38.2 (13.9)
Grade 6 group			250 (127)	35.3 (14.5)	250 (127)	41.1 (13.6)

Panel Data (3t): Maximum scale (full mark) of test score is 44

	Pre-baseline		Baseline		End-line	
	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)	N. of students (Female)	Mean score (SD)
Grade 4 group	240 (125)	17.2 (13.1)	240 (125)	20.7 (13.6)	240 (125)	30.8 (11.6)
Grade 5 group	220 (102)	23.5 (13.1)	220 (102)	28.0 (12.3)	220 (102)	32.7 (10.8)
Grade 6 group	250 (127)	28.1 (12.3)	250 (127)	30.0 (11.2)	250 (127)	34.7 (10.3)

Table 1-2: Number of students and test scores by grade group

Notes:
(1) "Grade 4 group" is defined as a group of students who were in grade 4 in the school year 2017-18. "Grade 5 group" and "Grade 6 group" are defined similarly.

(2) The first section "Pooled Data" is for students with test scores in 35 schools reported in Table 1-1, which include students who were not linked across different rounds of survey. Other three sections are for subsets of the first section. The second section "Panel Data (2t)" is for 1,068 students who were successfully linked between baseline and end-line. The third and fourth sections are for 710 students who were successfully linked between pre-baseline, baseline and end-line.

	(I) **4th / 5th** Pooled Model (1)	(II) **4th / 5th** Pooled Model (1-a)	(III) **4th / 5th** Panel: 2t Model (1-a)	(IV) **4th / 5th** Panel: 2t Model (2)
(Intercept)	32.829^{***}			
	(2.445)			
Grade 4 (Treatment)×Post	5.656^{***}	5.956^{***}	6.089^{***}	6.089^{***}
(λ)	(1.349)	(1.251)	(1.183)	(1.653)
Grade 4 (Treatment)	-8.310^{***}	-9.627^{***}	-7.854^{***}	
	(2.113)	(1.576)	(1.753)	
Post	6.697^{***}	6.493^{***}	6.247^{***}	6.247^{***}
	(1.002)	(1.028)	(0.863)	(1.206)
Female	-3.194^{**}	-3.280^{***}	-2.602^{**}	
	(1.175)	(0.896)	(1.170)	
Franco Arab school	-3.035			
	(3.339)			
Basic reading pilot school	-5.123			
	(3.707)			
School fixed effect (dummy)	No	Yes	Yes	No
Student fixed effect (dummy)	No	No	No	Yes
Adjusted R-Squared	0.151	0.865	0.880	0.958
Observations	1,995	1,995	1,384	1,384

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

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) Maximum scale (full mark) of math test score is 56.

Table 2-1: Average impact of PMAQ on math learning
(raw test score)

	(I)	(II)	(III)	(IV)
	4th / 5th	**4th / 5th**	**4th / 5th**	**4th / 5th**
	Pooled	Pooled	Panel: 2t	Panel: 2t
	Model (1)	Model (1-a)	Model (1-a)	Model (2)
(Intercept)	0.702^{***}			
	(0.159)			
Grade 4 (Treatment)× Post	0.367^{***}	0.387^{***}	0.387^{***}	0.387^{***}
(λ)	(0.088)	(0.081)	(0.075)	(0.105)
Grade 4 (Treatment)	-0.540^{***}	-0.625^{***}	-0.499^{***}	
	(0.137)	(0.102)	(0.111)	
Post	0.435^{***}	0.422^{***}	0.397^{***}	0.397^{***}
	(0.065)	(0.067)	(0.055)	(0.077)
Female	-0.207^{**}	-0.213^{***}	-0.165^{**}	
	(0.076)	(0.058)	(0.074)	
Franco Arab school	-0.197			
	(0.217)			
Basic reading pilot school	-0.333			
	(0.241)			
School fixed effect (dummy)	No	Yes	Yes	No
Student fixed effect (dummy)	No	No	No	Yes
Adjusted R-Squared	0.151	0.508	0.505	0.832
Observations	1,995	1,995	1,384	1,384

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

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) All the test scores at each round of survey are normalized by mean and standard deviation of baseline scores of 4th grade students.

Table 2-2: Average impact of PMAQ on math learning (standardized test score)

	(I)	(II)	(III)	(III)
	4th / 5th	**5th / 6th**	**5th / 6th**	**5th / 6th**
	Panel: 2t	Pooled	Panel: 2t	Panel: 2t
	Model (3)	Model (4)	Model (4)	Model (5)
Grade 4 (Treatment) × Pre-Treatment (Baseline) (λ')	-0.929			
	(1.768)			
Grade 5 × Post (λ'')		1.000	0.465	0.465
		(1.341)	(1.314)	(1.836)
Grade 5		-5.018**	-4.096	
		(1.991)	(2.502)	
Pre-Treatment (Baseline)	4.500***			
	(1.188)			
Post		5.709***	5.782***	5.782***
		(1.099)	(1.078)	(1.506)
Female		-2.295*	-2.244	
		(1.239)	(1.476)	
School fixed effect (dummy)	No	Yes	Yes	No
Student fixed effect (dummy)	Yes	No	No	Yes
Adjusted R-Squared	0.939	0.890	0.897	0.972
Observations	920	2,041	1,416	1,416

***p < 0.01, **p < 0.05, *p < 0.1

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) Maximum scale (full mark) of math test score in column (I) is 44. Maximum scale (full mark) of math test score in from column (II) to (IV) is 56.

(3) Column (I) of panel data is composed by pre-baseline and baseline data. Column (III) of panel data is composed by baseline and end-line data.

Table 3: Regression results of parallel trends test (raw test score)

	(I)	(II)	(III)
	4th / 5th	**4th / 5th**	**4th / 5th**
	Pooled	Pooled	Panel: 2t
	Model (1-b)	Model (1-c)	Model (1-c)
(Intercept)	32.582***		
	(2.472)		
Grade 4 (Treatment) × Post (λ)	5.761***	5.999***	6.139***
	(1.416)	(1.261)	(1.189)
Grade 4 (Treatment)	-8.359***	-9.650***	-7.879***
	(2.133)	(1.576)	(1.743)
Post	7.242***	6.773***	6.500***
	(1.098)	(1.080)	(0.919)
Female	-2.682**	-3.003***	-2.326*
	(1.247)	(0.964)	(1.195)
Franco Arab school	-2.908		
	(4.911)		
Basic reading pilot school	-4.979		
	(4.474)		
Post×Female	-1.100	-0.593	-0.552
	(0.955)	(0.639)	(0.655)
Post×Franco Arab school	-0.276		
	(3.813)		
Post×Basic reading pilot school	-0.257		
	(2.185)		
School fixed effect (dummy)	No	Yes	Yes
Student fixed effect (dummy)	No	No	No
Adjusted R-Squared	0.150	0.865	0.877
Observations	1,995	1,995	1,384

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) Maximum scale of math test score is 56.

Table 4-1: Regression results of robustness check (1) (raw test score)

	(I) **4th / 5th** Panel: 2t (Sub-sample) Raw Score Model (1-a)	(II) **4th / 5th** Panel: 2t (Sub-sample) Standardized Model (1-a)	(III) **4th / 5th** Panel: 2t (Sub-sample) Standardized Model (2)
Grade 4 (Treatment) × Post (λ)	4.648***	0.295***	0.288**
	(1.236)	(0.079)	(0.107)
Grade 4 (Treatment)	-2.823**	-0.179**	
	(1.187)	(0.075)	
Post	9.766***	0.621***	0.560***
	(0.793)	(0.050)	(0.069)
Female	-1.547	-0.098	
	(1.012)	(0.064)	
Post×Female	-2.045**	-0.130**	
	(1.003)	(0.064)	
School fixed effect (dummy)	Yes	Yes	No
Student fixed effect (dummy)	No	No	Yes
Adjusted R-Squared	0.909	0.458	0.703
Observations	672	672	672

***p < 0.01, **p < 0.05, *p < 0.1

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) Maximum scale (full mark) of math test score is 56. In column (B) and (C), test scores are standardized with mean and standard deviation of baseline scores of whole sample of 4th grade students.

(3) Panel data (baseline and end-line) is composed by students with baseline score from 10 to 37 score point (2nd and 3rd quartile of 4th grade student).

Table 4-2: Regression results of robustness check (2)

	(I)	(II)	(III)
	4th / 5th	**4th / 5th**	**4th / 5th**
	Panel: 2t	Panel: 2t	Panel: 2t
	(Sub-sample)	(Sub-sample)	(Sub-sample)
	Raw Score	Standardized	Standardized
	Model (1-a)	Model (1-a)	Model (2)
Grade 4 (Treatment) × Post (λ)	5.153***	0.328***	0.327**
	(1.380)	(0.088)	(0.123)
Grade 4 (Treatment)	-10.362***	-0.660***	
	(1.534)	(0.098)	
Post	7.043***	0.449***	0.438***
	(1.068)	(0.068)	(0.082)
Female	-3.741***	-0.238***	
	(1.237)	(0.079)	
Post×Female	-0.324	-0.021	
	(0.820)	(0.052)	
School fixed effect (dummy)	Yes	Yes	No
Student fixed effect (dummy)	No	No	Yes
Adjusted R-Squared	0.909	0.472	0.793
Observations	672	672	672

***p < 0.01, **p < 0.05, *p < 0.1

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) Maximum scale (full mark) of math test score is 56. In column (B) and (C), test scores are standardized with mean and standard deviation of baseline scores of whole samples of 4th grade students.

(3) Panel data (baseline and end-line) is composed by all of 4th grade students, and sub-sample of 5th grade students of schools which organized additional classes for the students.

Table 4-3: Regression results of robustness check (3)

		4th grade baseline test score				5th grade baseline test score			
		Higher	Med.	Lower	Total	Higher	Med.	Lower	Total
Sex	Male	43	86	34	163	43	89	48	180
	Female	47	93	57	197	30	84	38	152
	Total	90	179	91	360	73	173	86	332

Table 5: Number of 4th and 5th grade students by test score level and sex

	(I) **4th/5th** Panel: 2t Model (1-a)	(II) **4th/5th** Panel: 2t Model (6)	(III) **4th/5th** Panel: 2t Model (6)	(IV) **4th/5th** Panel: 2t Model (7)	(V) **4th/5th** Panel: 2t Model (8)
Grade4 (Treatment) × Post: DID (λ_0)	0.387*** (0.075)	0.134 (0.093)	0.341*** (0.082)	0.106 (0.106)	0.092 (0.127)
DID × Lowest score (λ_{1L})		0.585*** (0.123)		0.580*** (0.123)	0.493*** (0.166)
DID × Median score (λ_{1M})		0.211*** (0.070)		0.210*** (0.070)	0.269*** (0.096)
DID × Female (λ_2)			0.084 (0.050)	0.054 (0.056)	0.084 (0.089)
DID ×Lowest score ×Female (λ_{3L})					0.121 (0.173)
DID ×Median score ×Female (λ_{3M})					-0.111 (0.119)
Grade4 (Treatment)	-0.499*** (0.111)	-0.542*** (0.051)	-0.549*** (0.053)	-0.541*** (0.051)	-0.541*** (0.050)
Post	0.397*** (0.055)	0.397*** (0.055)	0.397*** (0.055)	0.397*** (0.055)	0.397*** (0.055)
Lowest score		-2.122*** (0.065)	-1.975*** (0.051)	-2.121*** (0.064)	-2.194*** (0.077)
Median score		-0.910*** (0.047)	-0.859*** (0.033)	-0.909*** (0.047)	-0.904*** (0.058)
Female	-0.165** (0.074)	-0.093* (0.050)	-0.112** (0.052)	-0.107** (0.050)	-0.147*** (0.053)
Lowest score×Female					0.157 (0.110)
Median score×Female					-0.003 (0.066)
School fixed effect	Yes	Yes	Yes	Yes	Yes
Student fixed effect	No	No	No	No	No
Adj. R^2	0.496	0.785	0.759	0.785	0.787
Observations	1,384	1,384	1,384	1,384	1,384

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) All the test scores are normalized with mean and standard deviation of baseline scores of 4th grade students. Maximum scale (full mark) of math test score is 56.

(3) Lowest score dummy takes 1 for students who obtained baseline score below 1st quartile score of 4th grade student baseline score. Median score dummy takes 1 for students who obtained baseline score more than 1st quartile score of 4th grade student baseline score, and below 3rd quartile of the score.

Table 6-1: Heterogeneous impacts of PMAQ on math learning (standardized test score)

	(I) **4th/5th** Panel: 2t Model (1-c)	(II) **4th/5th** Panel: 2t Model (6-a)	(III) **4th/5th** Panel: 2t Model (6-a)	(IV) **4th/5th** Panel: 2t Model (7-a)	(V) **4th/5th** Panel: 2t Model (8-a)
Grade4 (Treatment) × Post: DID (λ_0)	0.390*** (0.076)	0.219*** (0.070)	0.170** (0.068)	0.128 (0.077)	0.123 (0.090)
DID × Lowest score (λ_{1L})		0.272 (0.165)		0.268 (0.168)	0.198 (0.194)
DID × Median score (λ_{1M})		0.009 (0.077)		0.008 (0.080)	0.057 (0.114)
DID × Female (λ_2)			0.189** (0.093)	0.187* (0.094)	0.194** (0.095)
DID ×Lowest score ×Female (λ_{3L})					0.063 (0.396)
DID ×Median score ×Female (λ_{3M})					-0.086 (0.184)
Grade4 (Treatment)	-0.501*** (0.111)	-0.099** (0.042)	-0.104** (0.043)	-0.100** (0.042)	-0.101** (0.042)
Post	0.413*** (0.058)	0.137*** (0.045)	0.163*** (0.038)	0.182*** (0.047)	0.163*** (0.034)
Lowest score		-2.419*** (0.062)	-2.421*** (0.062)	-2.420*** (0.062)	-2.475*** (0.057)
Median score		-1.228*** (0.030)	-1.230*** (0.030)	-1.229*** (0.030)	-1.243*** (0.046)
Female	-0.148** (0.076)	-0.100*** (0.036)	-0.100*** (0.036)	-0.100*** (0.036)	-0.139*** (0.029)
Lowest score×Female					0.118* (0.066)
Median score×Female					0.036 (0.064)
Post × Lowest score		0.569*** (0.125)	0.744*** (0.097)	0.563*** (0.130)	0.555*** (0.153)
Post × Median score		0.481*** (0.061)	0.480*** (0.042)	0.482*** (0.059)	0.516*** (0.071)
Post×Female	-0.035 (0.042)	-0.052 (0.045)	-0.145* (0.074)	-0.149** (0.072)	-0.110 (0.075)
Post×Lowest score ×Female					0.049 (0.393)
Post×Median score ×Female					-0.075 (0.120)
School fixed effect	Yes	Yes	Yes	Yes	Yes
Student fixed effect	No	No	No	No	No
Adj. R^2	0.505	0.782	0.783	0.783	0.783
Observations	1,384	1,384	1,384	1,384	1,384

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

(1) Robust standard errors are clustered at school level, and in parenthesis. The variance of standard errors is adjusted by finite-sample correction.

(2) All the test scores are normalized with mean and standard deviation of baseline scores of 4th grade students. Maximum scale (full mark) of math test score is 56.

(3) Lowest score dummy takes 1 for students who obtained baseline score below 1st quartile score of 4th grade student baseline score. Median score dummy takes 1 for students who obtained baseline score more than 1st quartile score of 4th grade student baseline score, and below 3rd quartile of the score.

Table 6-2: Robustness check of heterogeneous impacts of PMAQ (standardized test score)