NBER WORKING PAPER SERIES

IMPROVING READING SKILLS BY ENCOURAGING CHILDREN TO READ IN SCHOOL: A RANDOMIZED EVALUATION OF THE SA AKLAT SISIKAT READING PROGRAM IN THE PHILIPPINES

Ama Baafra Abeberese
Todd J. Kumler
Leigh L. Linden

Working Paper 17185
http://www.nber.org/papers/w17185

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2011

We are indebted to many individuals who were significantly involved with the execution of this experiment. We wish to thank Catherine S. Alcaraz, Marie Angeles, Coly Los Baños, Clarissa Isabelle Delgado, Margarita L Delgado, Norlyn Gregorio, Elizabeth E. Zobel and all of the other staff members of the Sa Aklat Siskat Foundation for their support and assistance during the evaluation. All surveys were conducted by TNS Philippines. Finally, we are grateful to an anonymous donor for generously agreeing to fund this research effort. Without his or her help, this project would not have been possible. Leigh L. Linden is the corresponding author. Please direct all correspondence to leigh.linden@austin.utexas.edu.

Improving Reading Skills by Encouraging Children to Read in School: A Randomized Evaluation of the Sa Aklat Sisikat Reading Program in the Philippines
Ama Baafra Abeberese, Todd J. Kumler, and Leigh L. Linden
NBER Working Paper No. 17185
June 2011, Revised February 2013
JEL No. I21,I28,O15

ABSTRACT

We show that a short-term (31 day) reading program, designed to provide age-appropriate reading material, to train to teachers in their use, and to support teachers' initial efforts for about a month improves students' reading skills by 0.13 standard deviations. The effect is still present three months after the program but diminishes to 0.6 standard deviations, probably due to a reduced emphasis on reading after the program. We find that the program also encourages students to read more on their own at home, but find no evidence that improved reading ability improves test scores on other subjects.

Ama Baafra Abeberese
Department of Economics
Columbia University
420 West 118th Street, MC3308
New York, NY 10027
aba2114@columbia.edu

Todd J. Kumler
Department of Economics
Columbia University
420 West 118th Street, MC3308
New York, NY 10027
tjk2110@columbia.edu

Leigh L. Linden
Department of Economics
The University of Texas at Austin
2225 Speedway
BRB 1.116, C3100
Austin, Texas 78712
and NBER
leigh.linden@austin.utexas.edu

I. Introduction

Seven hundred and seventy-five million adults cannot read (UIS, 2011). A major factor is the poor quality of public schools in developing countries. Our understanding of the educational production function, however, is still limited. We know that providing resources to schools with no other inputs rarely improves overall student performance, and we know that if resources are paired with a larger array of inputs, students' test scores do improve (Glewwe and Kremer, 2006). Unfortunately, we do not know exactly what kind of additional support is necessary to make the additional resources effective. For reading in particular, a number of studies have demonstrated the effectiveness of large comprehensive changes. Banerjee et al. (2007), which studies an Indian remedial education program, is a good example. The intervention itself causes students' reading skills to improve, but because the intervention changes the educational environment along multiple dimensions – additional teachers, new pedagogical methods, new curriculum, changes to organization of the classroom, and additional resources – it is impossible to identify which components are necessary.

We approach this challenge by assessing the causal effects of a reading program that changes children's educational experience along a single dimension that is common to many more comprehensive reading programs: getting children to actively read age-appropriate books within the classroom. Schools in developing countries rarely encourage children to read. Curricula do not emphasize it, and most schools simply lack age appropriate reading material. As a result, comprehensive reading programs try to encourage children to read during the school day by providing age-appropriate reading material, segregating time during the school day for reading, and encouraging children through group readings, reading-based classroom games, and

other pedagogical changes focused on getting teachers to read books with their students.[2] From the perspective of understanding the effects of providing schools with additional resources, we assess a program that that provides teachers with new materials but also provides training in their use and periodic monitoring and support for about a month.

Using a randomized controlled trial set in Tarlac province of the Philippines, we analyze the causal impact of the Sa Aklat Sisikat (SAS) reading program which provides age-appropriate reading material for fourth grade classrooms, trains teachers to incorporating reading into their curriculum, and supports these changes through a 31-day reading marathon in which teachers encourage students to read while being directly supported by SAS. We randomly assigned, by school, 5,510 fourth-grade students in 100 schools to receive the intervention following a baseline assessment of students' reading skills at the start of the academic year and administered follow-up surveys after all of the marathons were complete (four months after baseline) and at the end of the academic year (seven months after baseline).

The results suggest that, in a short period of time, simply enabling and encouraging students to read age-appropriate books in school creates meaningful improvements in children's reading skills. On average, reading test scores increase by 0.13 standard deviations immediately after the end of the marathons. However, while the effects do persist, scores decline by 54 percent over the next three months. This suggests that providing resources and training alone is a viable short-term strategy for meaningfully improving children's reading skills, but by themselves they are insufficient to sustain those improvements.

[2] As part of larger programs, this might be combined with professional development for teachers, the creation of new infrastructure such as school libraries, student reading assessment techniques, changes in personal (such as the addition of a reading instruction coordinator or additional instructors), and often the use of new technologies that provide more functionality than traditional books (eReaders, tablets, or even computer assisted instruction).

The fade-out may result from teachers deemphasizing reading. During the marathons, the implementing NGO ensured that teachers provided time for reading, but while the teachers retained all of the materials after the program ended, they also regained control over the amount of time they dedicated to the subject. Consistent with this hypothesis, we find that the program increases the number of books children read in school in the last month by 7.17 during the marathon period, but by 56 percent less at the second follow-up. In fact, if we use the number of books read in the last month as a proxy for teachers' emphasis on in-school reading, the local average treatment effect (LATE) estimates of the change in standard deviations per book read is the same in both periods. This suggests that time spent on reading in-school was equally effective in both periods, but test scores declined because the amount of dedicated time declined after the first survey. To sustain long-term gains in reading skills, interventions like the read-a-thon may need to be paired with other program components, such as administrative and professional development interventions, aimed at encouraging a long-term focus on reading.

Finally, reading in school is often prioritized because of the hope that by making children better readers, they will be better equipped to learn other subjects and that by engaging with books, children will become interested in reading outside of school. We assess the first hypothesis by testing children in math and social studies, but we find no effect in either subject. We do, however, find that the in-school reading does encourage children to read more outside of school. For example, we find that children in the treatment group report reading 1.24 and 0.89 more books in the last month on the first and second follow-up surveys respectively.

The remainder of the paper is organized as follows. Section II provides an overview of the intervention. We describe the research design in Section III. Section IV documents the internal validity of the study, and in Section V, we estimate the effects of the treatment. We

compare the results to those of other studies of reading programs in Section VI. Finally, we conclude in Section VII.

II. The Sa Aklat Sisikat Read-A-Thon

The reading program evaluated in this study is a core program of Sa Aklat Sisikat,[3] a non-profit organization located in Manila with the stated goal of building a nation of readers. Since the inception of the organization in 1999, SAS has implemented its reading program in over 750 public schools in every province in the Philippines, reaching nearly 150,000 students. The program has three main components – providing schools with a set of age-appropriate books, training teacher to incorporate reading in the curriculum, and through a 31 day "read-a-thon", encourage children to read and support teachers as they incorporate reading into their classes. The program targets fourth grade students because the school system expects students to have developed sufficient reading fluency to enjoy reading independently by the fourth grade.[4]

Because most public schools lack age-appropriate reading material,[5] SAS donates 60 Filipino storybooks to every fourth-grade classroom at participating schools. The books are specifically selected for literary value as well as their potential appeal to students. Books are provided in both of the country's official languages, English and Filipino, so that teachers can use books that match their chosen language of instruction.[6]

[3] Sa Aklat Sisikat is loosely translated as "Books Make You Cool."
[4] Reading fluency is the point at which beginning readers rely less on the phonemic decoding to recognize individual words and begin to recognize whole words, allowing for an increase in reading speed and comprehension. Meyer and Felton (1999), for example, define fluency as "the ability to read connected text rapidly, smoothly, effortlessly, and automatically with little conscious attention to the mechanics of reading, such as decoding."
[5] For example, during our visits to local schools, we observed a few schools with libraries. However, most of the books were donated from developed countries. The subjects and writing styles were not age appropriate. It was thus not a surprise that teachers used them infrequently.
[6] For this reason, all evaluations in this study were conducted in Filipino. The Philippines has two official languages, Filipino and English, and under an existing executive order, schools are allowed to instruct students in either language. Books are provided in both languages, although in the sample, students were instructed in Filipino.

Prior to receiving the materials, fourth-grade teachers from each school attend a two-day training session in which they are taught to implement the read-a-thon and are given ideas for various reading lessons that incorporate reading in an engaging way. For 31 days after the training, they implement the read-a-thon. During this period, the students and teachers use the donated storybooks in hour-long daily reading activities including dramatic storytelling, literary games, and individual silent reading. Students are encouraged to read as many of the 60 story books as possible, and each student keeps track of the number of books she reads using an SAS supplied wall chart. Students also write their thoughts about the stories in reading notebooks. Finally, SAS also provides monitors who visit the schools to check that the program is being implemented correctly and to support teachers in using the new books.

While the read-a-thon itself only lasts 31 days, the partner schools keep the 60 storybooks following the completion of the program. The books are left for the class to use at their discretion, but the expectation is that the experiences created by the intensive read-a-thon will encourage teachers to continue using the books in their classes and students to continue reading.

III. Methodology

A. Research Design

The research sample consists of all fourth-grade classrooms at 100 elementary schools in Tarlac province in the Philippines. Prior to the experiment, Sa Aklat Sisikat had never conducted its reading program there.[7] In conjunction with the province superintendent, SAS selected nine geographically proximate districts within the province representing a broad range of academic performance levels. From these districts, 100 schools were chosen for the experiment; for most districts, all schools in the district were included in the study.

[7] In addition, relatively few other reading interventions had been conducted in the province.

A baseline survey was conducted in all 100 schools in July 2009. Following the survey, schools were assigned to the treatment and control groups using a matched-pair stratified randomization. Schools were divided into pairs within each district using the school level average baseline reading scores.[8] Within each pair, one school was assigned to the treatment group and the other the control group with equal probability. The read-a-thon was then implemented between the months of September and November.[9] Two follow-up surveys were conducted. The first was conducted immediately after the implementation of the read-a-thon in late November 2009 to measure the immediate effects of the intervention. The second was conducted at the end of the academic year in late February 2010 to determine whether the effects persisted after SAS ceased interacting with the treatment schools.

B. Data

Each survey round contained an assessment of children's reading skills. These exams were based in part on a national reading examination created and administered annually by the Philippine Department of Education.[10] The examination comprised sections covering six competencies. In the first part of the test (referred to as the "Written Test"), students are asked to silently read a written passage and answer written multiple-choice questions relating to the passage. Next, students were given one-on-one oral reading tests covering letter recognition, sound recognition,

[8] We have also estimated the primary specifications including fixed effects for the original groupings for the randomization. The results are consistent with those presented below. These results are available upon request.

[9] During the implementation of the read-a-thon, Tarlac experienced severe flooding that led to the cancellation of several days of school in many of the school districts. In addition, all-school events such as science fairs, town holidays, and standardized testing caused schools to take days off from the read-a-thon. However, all treatment schools completed the 31 day read-a-thon prior to the first follow-up examination.

[10] We chose to use sections of the national exam in order to ensure that both treatment and control groups were assessed using an instrument with which both groups were equally familiar. We wanted to avoid, for example, choosing an exam that might be geared towards the particular intervention being tested, and thus might favor the treatment students simply due to familiarity. The letter, sound, and word recognition sections were added to assess more basic competencies than typically tested on the official exam.

and word recognition. Finally, students were asked to read a passage aloud (referred to as the "Oral Reading" Test) and then to answer several questions about the passage orally ("Oral Reading Questions"). For each section, we normalized students' scores relative to the control distribution. Because the values for each section are not measured using the same units, we create a composite normalized reading score by averaging the normalized scores from each section and then normalizing the average, again relative to the distribution of the control group.

A local survey firm proctored and graded all of the examinations independently of the teachers to guarantee the validity of the test scores. In addition, teachers were not informed in advance of the content of the exam to prevent them from preparing students for the test. In order to ensure that a high percentage of students in the sample were tested, the survey team returned multiple times to many of the schools to test absent students.

Each survey also contained data unique to the individual round. In the baseline survey, we collected children's age, gender, height, weight, number of siblings, religion, and the dialect spoken at home. In the follow-up surveys, we collected information on children's reading habits as well as tested students in other subjects to investigate possible spillovers from the intervention. The reading survey asked students about the number of books they read in the last week and the last month both in and out of school. In addition, we asked students to name the title and to describe the plot of the last book they read to assess the validity of their responses. For the alternate subject tests, we tested a different subject each round. In the first follow-up survey, we tested children's math skills, and in the second one, we tested children's knowledge of social studies, the most reading intensive alternate subject.

C. Statistical Models

W utilize three basic models. First, we employ a simple difference specification to directly compare the treatment and control groups:

$$Y_{is} = \alpha + \beta_1 T_s + \varepsilon_{is} \quad (1)$$

where Y_{is} is the outcome of interest for child i in school s; and T_s is an indicator variable for whether the school received the reading program. Hence, the estimate of the coefficient β_1 indicates the differences between treatment and control schools. We utilize this model to compare baseline differences in socio-demographic characteristics and test scores and to estimate the effect of the reading program on follow-up test scores and reading habits.

Since the reading program is randomly assigned to schools and therefore independent of baseline characteristics, inclusion of observable baseline characteristics and baseline test scores as control variables in equation (1) improves the precision of the estimated treatment effect. We also run the following specification:

$$Y_{is} = \alpha + \beta_1 T_s + \beta_2 X_{is} + \omega_d + \varepsilon_{is} \quad (2)$$

where Y_{is} and T_s are defined as in equation (1), and where X_{is} is a vector of baseline student characteristics including composite baseline reading test score, gender, age, religion dummies, dialect dummies, and body mass index (BMI). Since the randomization was stratified within district, we also include district fixed effects, ω_d, in equation (2).

Finally, we test the validity of the experiment by comparing the effect of the treatment on the relative characteristics of the children who attrit from the sample between the baseline survey and the two follow-up surveys. We run the following difference in differences model:

$$Y_{is} = \alpha + \beta_1 T_s + \beta_2 Attrit_{is} + \beta_3 T_s * Attrit_{is} + \varepsilon_{is} \quad (3)$$

The variables Y_{is} and T_s are defined as before, and $Attrit_{is}$ is an indicator variable equal to one if student i enrolled in school s is not present in the follow-up data. The estimate of β_2 then provides the average differences between attritors and non-attritors in the control group, and the estimate of β_3 captures the difference-in-differences between attritors and non-attritors in the treatment and control groups.

Because outcomes may be correlated within school, failure to correct the standard errors could result in an overestimate of the precision of the treatment effects (Bertrand, Duflo, Mullainathan, 2004). We therefore cluster the standard errors at the school level (the level of randomization) in all of the above models.

IV. Internal Validity

Randomly assigning schools to the intervention ensures that assignment is orthogonal to student characteristics that may be correlated with the outcomes of interest. If this holds, then any differences in outcomes between the two groups post-intervention can be causally attributed to the intervention. To check that student characteristics in each group were indeed similar, we run regressions of student characteristics from the baseline survey on treatment assignment, and then we verify that any changes in the sample due to attrition are also uncorrelated with treatment assignment.

We present the comparison of students at baseline in Table 1. Column 1 contains the average characteristics for the control group. Columns 2 and 3 present the estimated differences between the treatment and control groups. The results in column 2 do not include any controls, while those in column 3 control for district fixed effects. Panels A and B contain standardized reading test scores and demographic characteristics, respectively.

The differences in average characteristics between the control and treatment groups are all practically small and mostly statistically insignificant. In Panel A, none of the differences in test scores are statistically significant. Figure 1 shows a plot of the distribution of the standardized overall reading test score for the treatment group (solid line) and the control group (dashed line). These distributions almost overlap completely, further corroborating the comparability of the research groups. In Panel B, the only demographic variables with statistically significant differences are those related to religion, but these differences are small in magnitude. For instance, 74 percent of students in the control group are Catholics compared to 69 percent in the treatment group, yielding a minimal difference of 5 percentage points. The randomization thus appears to have successfully created similar treatment and control groups.

Although the baseline comparisons presented in Table 1 and Figure 1 show that the treatment and control groups were similar at baseline, it is possible that non-random attrition from the two groups between the baseline and follow-up surveys may have rendered the two groups incomparable. Table 2 shows the attrition rates for both groups and the differences between the two. There are no statistically significant differences between the attrition rates for the control and treatment groups. For both groups, approximately 5 percent of the students who were tested during the baseline survey were absent during the first follow-up survey, and 11 percent were absent during the second survey. Comparing the rates across research groups, the rates are the same in the first follow-up and differ by only 2 percentage points in the second (10 percentage points in the treatment schools and 12 in the control).

Columns 4 through 6 provide estimates of the attrition rates between follow-up surveys. Overall, 86 percent of the students were present at both follow-up surveys (column 4), and the difference in the rates between research groups is small. Similarly, 91 percent of students who

were present at follow-up one were also present at the second follow-up, and of those present at the second, 97 percent were present at the first.

Even though the attrition rates were similar for both groups, the characteristics of the attritors and non-attritors could have still differed between the two groups. We check this possibility in Table 3 for the first follow-up survey. The results for the second follow-up survey are similar and presented in the Appendix (Table A1). Panel A focuses on test scores while Panel B focuses on demographic characteristics. Columns 1 and 2 contain the average characteristic for non-attritors in the control and treatment groups, respectively, while column 3 contains the difference between these averages estimated using equation (1). All of the differences are statistically insignificant with the exception of the proportion of non-attritors who are Catholic. However, this difference is small in magnitude (5 percentage points) and is identical to the difference found for the entire sample during the baseline survey.

The last three columns of Table 3 show that the differences between the characteristics of the non-attritors and attritors are similar across the two groups, indicating that there was no selection in the sample due to attrition. Column 4 presents the difference in average characteristic between the non-attritors and the attritors in the control group. Column 5 presents the same statistic for the treatment group, and column 6 presents the difference between the two statistics using equation (3). These differences are mostly statistically insignificant, and all of them are small in magnitude. We therefore conclude that the comparability of the control and treatment groups was sustained throughout the follow-up surveys.

V. Results

A. Effect on Reading Habits

The primary goal of the SAS reading program is to provide children the opportunity and means to read in-school and to encourage them to do so. As a result, we start by assessing whether or not students in schools assigned to the program did, in fact, read more in school. Table 4 compares reading rates across the two groups based on survey responses during the first and second follow-up surveys. Variables included in the table are students' responses to questions on whether or not they had read a book and the number of books read in the last week and month. To check that students who claim to have read a book actually did so, we also recorded whether or not children could name and/or summarize the last book they read.

The first three columns report results from the first follow-up survey, while the last three columns report results from the second follow-up survey. For each survey, the first column provides the average responses for the control group. The second and third columns provide estimates of the differences between groups without controls (equation (1)) and with controls (equation (2)), respectively.

During the period in which the read-a-thon was implemented, the program did significantly increase the amount of reading students did in school. The results in columns 1 and 3 show that 68 percent of the students in the control group reported reading a book in school in the past week on the first follow-up survey, and the program increased this by 19 percentage points. The students in the control group reported reading an average of 1.9 books in school in the past week and the program increases this by 2.3 books. In the past month, the program increased the number of books read by 7.2 books.

Further corroborating these results,[11] we find significant differences in the propensity to read if we only consider a child as having read a book if he or she can provide specific information about the last book read. If we consider children to have read a book only if they claim to have read a book and can provide the title, 53 percent of students in the control group read a book in the last week and the increase due to the program was 30 percentage points. If the condition is to describe the plot, the program caused 23 percentage points more children to have read a book. All of these results are statistically significant at the one percent level and are basically the same for the different specifications presented in columns 2 and 3.

After the program, the effects on student reading seem to have continued, but at about half of the previous rate. In terms of the probability that a student read a book (row one) or could identify the title (row four) or plot (row five), the effects of the program seem to be the same as during the read-a-thon period. However, when the questions focus on the number of books rather than just whether or not a child has read any book, the magnitudes decline. The effect on the number of books read in the last week is a statistically insignificant 0.86 and the effect on the number of books read in the last month is 3.12, statistically significant at the one percent level. This suggests that the program did have a long-term effect, but that the amount of time children spend reading declines after the direct support of the program is removed.

[11] One of the concerns with these self-reported numbers was that, knowing that they are generally expected to read, students might lie to surveyors about having read a book recently. The additional questions about the books provide one check. Also interesting in this respect, is the stability of the estimates for the fraction of children having reported reading a book (and being able to provide title and description) across the various surveys. For the control students, for example, the largest differences in rates is for the fraction of students reporting reading a book and being able to describe the book in Panel A at 9 percentage points. The next largest difference is 6 percentage points (being able to give the title and reporting having read a book in Panel A). The other five differences between the surveys are all in the range of 2-3 percentage points.

B. Effect on Reading Ability

We now explore the extent to which the changes in reading affected students' reading ability. Table 5 presents estimates of the differences between the standardized average reading test scores of the control and treatment groups. We present three estimates: an estimate of the treatment effect without any controls (column one, equation (1)), an estimate including only demographic characteristics (column two), and an estimate controlling for demographic characteristics and district fixed effects (column three, equation (2)).

Starting with the results from the first follow-up survey, the program had a distinct immediate effect on students' reading skills of 0.13 standard deviations. The results are consistent across the various specifications, highlighting the comparability of the treatment and control groups. And, in our preferred specification (column three), the results are statistically significant at the one percent level. Consistent with the reduction in the amount of reading children do at school, we find that the treatment effect declines between the first and second follow-up surveys to 0.06. The estimate is still consistent across the specifications and statistically significant at the five percent level, but it is 54 percent smaller.

To further investigate this relationship, we use the number of books a child reports reading in the last month in school as a proxy for the time teachers spend on reading. We then estimate local average treatment effects of reading on students reading test scores.[12] If the decline in test scores did result from the reduction in the time teachers spent on reading, then the coefficient on the LATE estimate should be similar for both surveys. This is, in fact, the case. The estimates are 0.0175 (p-value 0.017) and 0.0199 standard deviations per book (p-value

[12] It is important to note that this cannot be interpreted as the causal effect of reading a book in school on test scores because reading in the last month is almost certainly correlated with other activities, such as number of books read in the previous month. However, these same correlations make it good proxy for reading emphasis.

0.056) for the first and second surveys respectively.[13] This suggests that the effect of the curriculum change remained consistent across the two periods and that the decline in test scores is due to the reduced focus on children reading after the read-a-thon period.

We also investigated differences in the observed treatment effects for a number of subsets of our sample defined through the baseline survey. In results not presented in this manuscript,[14] we test for differences in treatment effects by gender, age, language spoken at home, and baseline reading score. We find almost no evidence of systematically different treatment effects for different types of students for either follow-up period. The one exception is that we find that, for the first follow-up period, the treatment effect increases with students' baseline test score. In a regression interacting treatment effect with baseline score, we find that students experience a 0.12 standard deviation increase at the control baseline mean (statistically significant at the one percent level) and then experience an increased effect of 0.09 standard deviations for each additional standard deviation they scored at baseline (significant at the ten percent level). While both coefficients are still positive at the second follow-up, the magnitudes are much smaller. They are 0.06 for the effect at the control mean and 0.04 for the interaction effect, and only the first remains statistically significant at conventional levels (five percent level). It may be that stronger students are more able to utilize the supplied books independently of the teacher than weaker students, and as a result, the benefit more than their classmates from the opportunity to read in school. This is consistent with other studies that have observed that stronger students tend to experience larger treatment effects from self-directed interventions (e.g. He, MacLeod, and Linden, 2008).

[13] We perform the same estimates using the number of books read in the last week and find similar results. Although the estimates are less precise, possibly because the number of books read in the last week is a weaker proxy than the number read in the last month.
[14] Results are available upon request.

Finally, Table 6 disaggregates the effect on the reading test by competency. The first two columns report the results from the first follow-up survey, while the last two columns report the results from the second. The first column for each of the follow-up surveys provides the differences between the average test scores of the control and treatment groups using equation (1). The second column provides the estimates with the full set of controls using equation (2). Finally, the last two rows of Table 6 report the chi-squared statistic and p-value from a test of the joint significance of the coefficients on the treatment indicator from the regressions for the individual components of the reading test, estimated using seemingly unrelated regressions.

On both the first and second follow-up surveys, the program had a positive effect on most of the components on the test (the written part being the one exception). In the first follow-up survey, there are sizable treatment effects on sound recognition, word recognition, and the ability to answer questions from the oral reading passage. The effects on word recognition and the oral questions are individually statistically significant at the one and ten percent levels respectively, but the effect on sound recognition is not statistically significant at conventional levels. The joint test of an effect on all skills shows that the results for the different components are jointly statistically significant with a p-value of 0.013.

In the second follow-up, the results are generally the same, but as with the overall average diminished in magnitude. In this round, word recognition and oral reading effects are individually statistically significant with word recognition remaining the strongest effect of the program. But again, all components show positive treatment effects except the written test and sound recognition. Overall, the joint hypothesis of the significance of the effects on each component is still statistically significant at the one percent level.

C. Other Outcomes

In addition to a direct effect on students' reading abilities, encouraging children to read in school could also have effects on other outcomes as well. First, due to their improved reading skills, students may be better equipped to study other subjects since an improved ability to read might allow them to read textbooks or other classroom materials. Table 7, which has as similar format as Table 5, presents the estimated treatment effects on the other subjects tested during the follow-up surveys. We find no effect on either subject. Although in the second follow-up survey, the treatment effect for social studies is of the same magnitude as the effect on the reading test, but measured less precisely.[15]

Second, although teachers reportedly never allowed children to take books home,[16] reading more in school may also make children more likely to read outside of school. To assess this, we asked children the same kinds of questions about their reading activities at home as at school. Table 8 presents the estimated results. Overall, we find a small, but statistically significant effect on both surveys. For the first follow-up survey, only the 1.24 effect on the number of books read in the last month is individually significant, but jointly, all of the estimates are statistically significant at the 1 percent level. In the second follow-up survey, the standard errors fall significantly, and although the point estimates are similar to those in the first round, all but the effect on the probability of reading any book outside of schools (row one) are statistically significant at conventional levels. This suggests that engaging students in reading in school

[15] This may be due to the fact that the social studies test was much less comprehensive than the reading test.
[16] Teachers do not allow children to take the books home to read for fear that the children might lose them. Thus, for children to read at home, they must have accessed books independently of the intervention.

increases the number of books read outside of school as well as in school. But like the effect on test scores, the effect declined after the read-a-thon. [17]

VI. Comparisons to Other Program

At 0.13 standard deviations, the direct effect of the program during implementation is consistent with the effects of many programs on native language skills.[18,19] Machin and McNally (2008), Muralidharan and Sundararaman (2011), Muralidharan and Sundararaman (2010), and Das et al. (2010) find effects in this range, for example. This is, however, smaller than the effects observed for others. Banerjee et al. (2007) find a total effect of 0.187 standard deviations for a remedial education program immediately after two years of implementation, and He, Linden, and MacLeod (2009) find a one year effect of 0.695 standard deviations for a pre-school reading program after a year of implementation.

One can also take into account the costs of the program. However, it is important to keep two caveats in mind. First, this program only affects reading skills while many of the other programs are designed to affect directly multiple subjects. Second, the few studies report treatment effects for native language skills as well as information on program costs. With these

[17] Verifying that the effect of time spent reading in school is consistent for both periods, as they are for test scores, is complicated by the imprecision of the estimated effect on reading outside of school on the first follow-up survey. We find consistent results for the number of books read in the last month, but for the number read in the last week, the 95 percent confidence interval around the first-year follow-up LATE estimate, (-0.192, 0.903), includes estimates that are more than twice the magnitude of those estimated using the second follow-up survey.

[18] These comparisons require two important caveats. First, we restrict the set of programs to only those that attempted to improve students' native language skills because the effects of these programs are usually much lower than effects observed for other subjects, such as math and foreign languages (see Banerjee et al. (2007) for example). Second, we focus on the effect of the program observed directly following program implementation because these are the effects most often reported. While fade-out after program implementation is an important issue, few studies estimate such longer-term treatment effects, and those that due, typically find significant fade-out. For example, Banerjee et al. (2007) find that, while the effects of the remedial education program on reading persist for the weakest students, the 0.33 standard deviation overall treatment effect of a remedial education program after two years of treatment declines to a statistically insignificant 0.040 standard deviations a year after students leave the intervention.

[19] The effect is, of course, larger than those of programs that are found to have no overall average effect such as Fyer (2011) and Kim and Guryan (2010).

caveats in mind, the SAS reading program seems to fall in the mid-range of programs adjusting for cost. The cost per tenth of a standard deviation gain per child is 8.52 USD.[20] This is higher than the interventions considered by Banerjee et al (2007), Muralidharan and Sundararaman (2011), Muralidharan and Sundararaman (2010), and Das et al. (2011), all of which cost between 1.53 USD and 3.11 USD per tenth a standard deviation per child for native language skills in India. It is, however, much lower than Machin and McNally (2009) whose UK-based program costs $46.42 USD per tenth of a standard deviation per child.[21]

VI. Conclusion

We demonstrate that a short-term reading program that provides age-appropriate reading material and trains teachers to use it can have a significant effect on the reading ability of primary school children. Reading test scores of students increased by 0.13 of a standard deviation immediately following the intervention. These gains in reading ability were still evident, albeit smaller at about 0.06 standard deviations, three months after the end of the intervention. We also find that the additional focus on reading in school causes a small increase in the number of books children read on their own at home, but we find no evidence that improved reading skills translate into better performance in other subjects

These results suggest that providing additional resources along with training and support in their use can improve students test scores in the short-run, but in order to sustain these gains, additional support is necessary. We show that while the focus on reading in the curriculum diminishes between the first and second surveys, the effectiveness of reading does not change.

[20] All cost information has been adjusted to 2010 USD using the US Consumer Price Index for All Urban Consumers including all available items.

[21] India, Philippines, and the UK are, of course, very different countries, making it difficult to rank programs based solely on this measure of cost-effectiveness. However, since there are so few examples of successful reading programs that report such cost information, we present those that we have found.

The decline in test scores seems to be solely due to a reduction in the emphasis on reading in the curriculum after the direct support of the NGO is removed. This suggests that teachers retain the ability to teach reading more effectively, but simply choose to do so less often. To make the additional resources effective, researchers may need to identify additional supports that are capable of sustaining teachers' focus on reading. This might, for example, be a change in the incentives faced by teachers (for example, having principals change the reward structure faced by teachers to emphasize reading) or strategies for reminding teachers of the importance of allocating time for reading (such as text messages or even long-term periodic monitoring).

VII. Bibliography

Baker, S., Gersten, R. and Keating, T. (2000) "When Less May Be More: A 2-Year Longitudinal Evaluation of a Volunteer Tutoring Program Requiring Minimal Training," *Reading Research Quarterly.* 35(4): 494-519.

Banerjee, A., Cole, S., Duflo, E. and Linden, L. (2007) "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics.* 122(3): 1235-1264.

Bertrand, M., Duflo, E. and Mullainathan, S. (2004) "How Much Should We Trust Difference-in-Differences Estimates?," *Quarterly Journal of Economics.* 119(1): 249–275.

Borkum, E., He, F. and Linden, L. (2009). "School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program," Working paper. Department of Economics. The University of Texas at Austin.

Center, Y., Wheldall, K., Freeman, L., Outhred, L. and McNaught, M. (1995) "An Evaluation of Reading Recovery," *Reading Research Quarterly*. 30(2): 240-263.

Das, Jishnu et. Al (2011) "School Inputs, Household Substitution, and Test Scores," Working Paper. Department of Economics. The University of California at San Diego.

Fryer, Roland (2011) "Financial Incentives and Student Achievement: Evidence from Randomized Trials," *Quarterly Journal of Economics*. 126(4): 1755-1798.

Glewwe, P. and Kremer, M. (2006) "Schools, Teachers and Education Outcomes in Developing Countries," *in* Hanushek, A. and Welch, F. (eds.): *Handbook of the Economics of Education* (vol. 2). Amsterdam: Elsevier.

Glewwe, P., Kremer, M., Moulin, S. and Zitzewitz, E. (2004) "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics*. 74:251-286.

Glewwe, P., Kremer, M. and Moulin, S. (2009) "Many Children Left Behind? Textbooks and Test Scores in Kenya," *American Economic Journal: Applied Economics*. 1(1): 112-135.

Hanushek, E. (1995) "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer*. 10(2): 227-246.

Hanushek, E. (2006) "School Resources," *in* Hanushek, A. and Welch, F. (eds.): *Handbook of the Economics of Education* (vol. 2) Amsterdam: Elsevier.

Hanushek, E. and Wossman, L. (2007) *The Role of Education Quality for Economic Growth.* Washington, DC: World Bank.

He, F., Linden, L. and MacLeod, M. (2008) "Teaching What Teachers Don't Know: An Assessment of the Pratham English Language Program," Working Paper. Department of Economics. The University of Texas at Austin.

He, F., Linden, L. and MacLeod, M. (2009) "Teaching Pre-Schoolers to Read: A Randomized Evaluation of the Pratham Shishuvachan Program," Working Paper. Department of Economics. The University of Texas at Austin.

Kim, B.S. (2007) "The Effects of a Voluntary Summer Reading Intervention on Reading Activities and Reading Achievement," *Journal of Educational Psychology***.** 99(3): 505–515.

Kim, J.S. and J. Guryan (2010) "The Efficacy of a Voluntary Summer Book Reading Intervention for Low-Income Latino Children from Language Minority Families," *Journal of Educational Psychology.* 102(1): 20-31.

Machin, S. and McNally, S. (2009) "The Literacy Hour," *Journal of Public Economics*. 92:1441-62.

Meyer, Marianne S. and Rebecca H Felton. (1999) "Repeating Reading to Enhance Fluency: Old Approaches and New Directions," *Annals of Dyslexia.* 49:283-306.

Muralidharan, Karthik and Venkatesh Sundararaman (2010) "Contract Teachers: Experimental Evidence from India," Working Paper. Department of Economics. The University of California at San Diego.

Muralidharan, Karthik and Venkatesh Sundararaman (2011) "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 119(1): 39-77

Rouse, C.E. and Krueger, A.B. (2004) "Putting Computerized Instruction To The Test: A Randomized Evaluation of a `Scientifically-Based' Reading Program," *Economics of Education Review*. 23(4): 323-338; 2004.

Scarborough, H.S. (2001) "Connecting Early Language and Literacy to Later Reading (Dis)Abilities: Evidence, Theory, and Practice," In S.B. Neuman and D.K. Dickinson (Eds.), *Handbook of Early Literacy Research (Volume 1)* (pp. 97-110). New York, NY: Guilford Press.

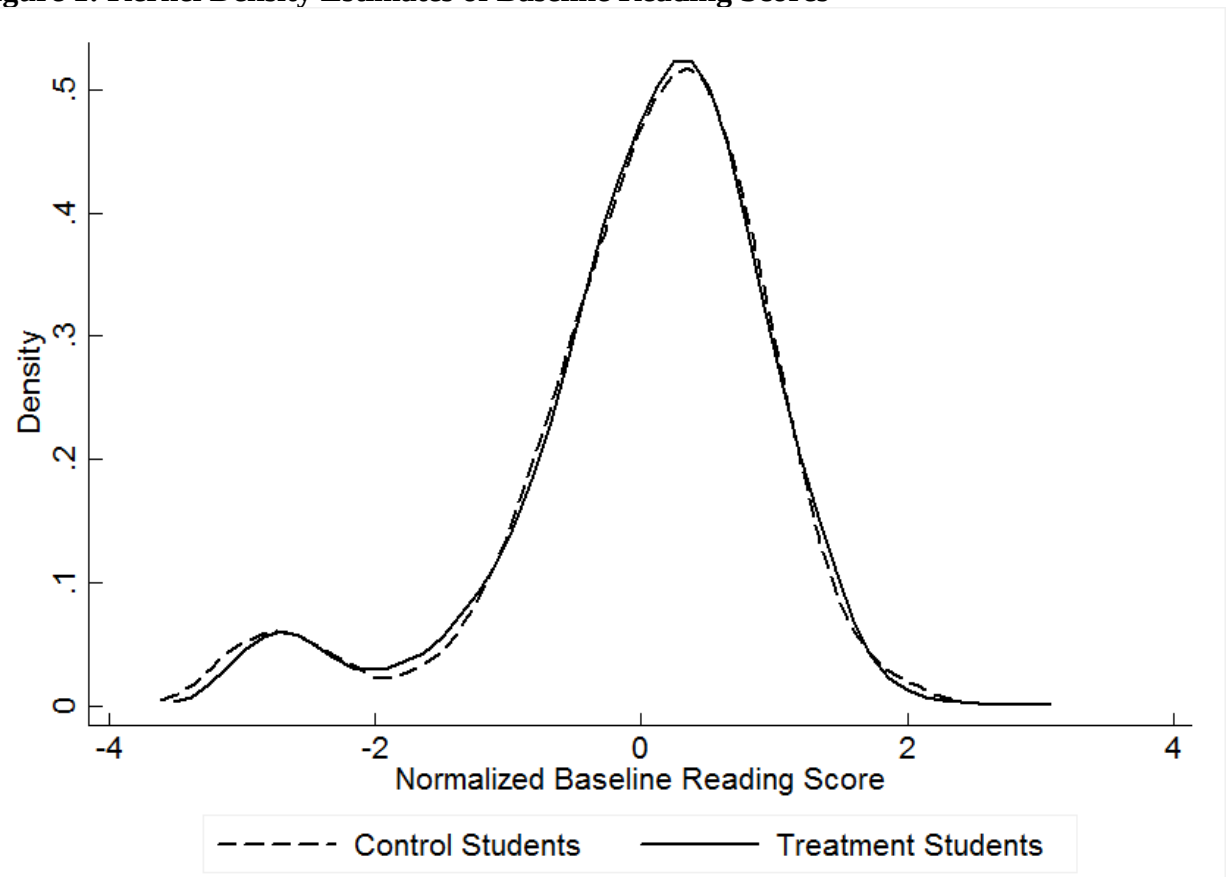
Shanahan, T. and Barr, R. (1995) "Reading Recovery: An Independent Evaluation of the Effects of an Early Instructional Intervention for At-Risk Learners," *Reading Research Quarterly*. 30(4): 958-996.

Vellutino, F.R. and Scanlon, D.M. (2001) "Emergent literacy skills, early instruction, and individual differences as determinants of difficulties in learning to read: The case for early intervention," In S.B. Neuman and D.K. Dickinson (Eds.), *Handbook of Early Literacy Research (Volume 1)* (pp. 295-321). New York, NY: Guilford Press.

Wasik, B.A. (1998) "Volunteer Tutoring Programs in Reading: A Review," *Reading Research Quarterly.* 33(3): 266-291.

Wasik, B.A. and Slavin, E. (1993) "Preventing Early Reading Failure with One-to-One Tutoring: A Review of Five Programs," *Reading Research Quarterly*. 28(2): 178-200.

Figure 1: Kernel Density Estimates of Baseline Reading Scores



Note: This figure presents kernel density estimates of the baseline total normalized reading score distributions for the treatment and control groups. Distributions estimated using an Epanechnikov kernel with a bandwidth of 0.2 standard deviations.

Table 1: Baseline Comparison

Dependent Variable	Control Mean (1)	Treatment Difference No Controls (2)	Treatment Difference District FE (3)
Panel A: Standardized Baseline Reading Test Scores			
Written Test	0.00	0.02	0.03
		(0.05)	(0.04)
Letter Recognition	0.00	0.03	0.03
		(0.04)	(0.04)
Sound Recognition	0.00	-0.10	-0.06
		(0.07)	(0.07)
Word Recognition	0.00	0.02	0.03
		(0.06)	(0.06)
Oral Reading	0.00	0.02	0.04
		(0.06)	(0.06)
Oral Reading Questions	0.00	0.04	0.05
		(0.07)	(0.05)
Average Score	0.00	0.01	0.03
		(0.06)	(0.06)
Panel B: Individual Characteristics			
Age	9.37	-0.01	-0.01
		(0.05)	(0.05)
Female	0.48	-0.01	-0.01
		(0.01)	(0.01)
Height	128.44	-0.05	-0.05
		(0.32)	(0.26)
Weight	56.56	0.83	0.57
		(0.77)	(0.69)
BMI	15.42	0.23	0.15
		(0.17)	(0.16)
Siblings	3.88	0.07	0.08
		(0.11)	(0.09)
Catholic	0.74	-0.05*	-0.05**
		(0.03)	(0.02)
INC	0.13	0.01	0.01
		(0.01)	(0.01)
Aglipayan	0.02	0.00	0.01
		(0.01)	(0.01)
Born Again	0.06	0.02	0.02*
		(0.01)	(0.01)
Protestant	0.02	0.01	0.01
		(0.01)	(0.01)
Other Religion	0.02	0.01	0.00
		(0.01)	(0.00)
Filipino	0.44	0.01	0.01
		(0.07)	(0.04)
Iloco	0.19	-0.05	-0.03
		(0.05)	(0.03)
Kapampangan	0.37	0.04	0.02
		(0.09)	(0.03)
Pangasinan	0.01	< 0.01	< 0.01
		(< 0.01)	(< 0.01)
Other Language	< 0.01	< 0.01	< 0.01
		(< 0.01)	(< 0.01)
Observations	2596	5510	5510

Note: This table presents a comparison of students who took the baseline survey in the control and treatment schools. Column 1 contains the average characteristic of the students in the control schools. Columns 2 and 3 contain estimates of the average difference in characteristics between the control and treatment students, without any controls and with only district fixed effects, respectively. Panel A contains students' standardized baseline test scores, and Panel B contains students' demographic characteristics. Standard errors are clustered by school. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 2: Attrition Rates

		Of Baseline Students:				
	No. of Students at Baseline	**Fraction at Follow-Up One**	**Fraction at Follow-Up Two**	**Fraction at Follow-Up One and Two**	**Fraction from Follow-Up One at Follow-Up Two**	**Fraction from Follow-Up Two at Follow-Up One**
	(1)	(2)	(3)	(4)	(5)	(6)
Control Schools	2596	0.95	0.88	0.86	0.90	0.97
Treatment Schools	2914	0.95	0.90	0.87	0.92	0.97
Difference		< 0.01	0.02	0.02	0.02	0.00
		(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Total	5510	0.95	0.89	0.86	0.91	0.97

Note: This table shows the fraction of baseline students who took the first and second follow-up surveys from the control and treatment schools and an estimate of the difference between the two groups. Column 1 contains the number of students who took the baseline survey. Column 2 shows the fraction of baseline students who took the first follow-up survey, while column 3 shows the fraction of baseline students who took the second follow-up survey. Column 4 contains the fraction of baseline students who took both the first and second follow-up surveys. Column 5 contains the fraction of baseline students at the first follow-up survey who took the second follow-up survey. Column 6 contains the fraction of baseline students at the second follow-up survey who took the first follow-up survey. The estimated differences in the third row are estimated using equation (1). Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 3: Attrition Patterns, Follow-Up One

	Non-Attritors			Non-Attritors Less Attritors		
	Control Mean	Treatment Mean	Treatment Difference	Control Difference	Treatment Difference	Difference-in-Difference
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Standardized Baseline Reading Test Scores						
Written Test	0.00	0.03	0.02	0.10	0.08	-0.01
			(0.05)			(0.12)
Letter Recognition	0.00	0.04	0.04	0.09	0.16	0.07
			(0.04)			(0.11)
Sound Recognition	0.01	-0.09	-0.10	0.22	0.11	-0.11
			(0.07)			(0.12)
Word Recognition	0.01	0.05	0.04	0.25	0.47	0.22
			(0.06)			(0.18)
Oral Reading	0.01	0.04	0.03	0.21	0.43	0.22
			(0.06)			(0.21)
Oral Reading Questions	0.00	0.05	0.05	0.06	0.19	0.13
			(0.07)			(0.14)
Average Score	0.01	0.03	0.02	0.22	0.34	0.12
			(0.06)			(0.16)
Panel B: Individual Characteristics						
Age	9.32	9.31	-0.01	-0.91	-0.90	0.00
			(0.04)			(0.29)
Female	0.48	0.47	-0.01	0.07	0.10	0.04
			(0.01)			(0.06)
Height	128.27	128.27	0.00	-3.34	-2.33	1.00
			(0.30)			(1.57)
Weight	56.35	57.25	0.90	-4.04	-2.68	1.36
			(0.79)			(2.13)
BMI	15.41	15.64	0.23	-0.21	-0.07	0.14
			(0.18)			(0.32)
Siblings	3.84	3.92	0.08	-0.83	-0.65	0.17
			(0.11)			(0.25)
Catholic	0.74	0.69	-0.05*	0.10	0.09	-0.01
			(0.03)			(0.07)
INC	0.12	0.13	0.01	-0.08	-0.05	0.03
			(0.01)			(0.04)
Aglipayan	0.02	0.03	0.00	0.01	0.00	-0.01
			(0.01)			(0.02)
Born Again	0.06	0.08	0.02	0.00	-0.05	-0.04
			(0.01)			(0.05)
Protestant	0.03	0.04	0.01	0.02	0.00	-0.02
			(0.01)			(0.02)
Other Religion	0.02	0.02	0.01	-0.04	0.01	0.05**
			(0.01)			(0.02)
Filipino	0.44	0.44	0.01	-0.04	-0.01	0.02
			(0.07)			(0.07)
Iloco	0.19	0.15	-0.04	-0.03	0.06	0.09*
			(0.05)			(0.05)
Kapampangan	0.37	0.40	0.03	0.06	-0.02	-0.08
			(0.09)			(0.06)
Pangasinan	0.01	< 0.01	< 0.01	0.01	-0.01	-0.02*
			(< 0.01)			(0.01)
Other Language	< 0.01	< 0.01	< 0.01	< 0.01	-0.02	-0.02
			(< 0.01)			(0.02)
Observations	2463	2765	5228	2596	2914	5510

Note: This table presents a comparison of the attrition rates bewteen the treatment and control groups for the first follow-up survey. Column 1 contains the average characteristics of the baseline control students who took the first follow-up survey, while column 2 contains the average characteristic of the baseline treatment students who took the first follow-up survey. Column 3 presents an estimate of the average difference in characteristics between the baseline control and treatment students who took the first follow-up survey. Columns 4 and 5 present the average difference in characteristics between the baseline students who took the first follow-up survey (non-attritors) and those who did not (attritors) for the control and treatment groups, respectively. Column 6 contains an estimate of the difference between the average differences in columns 4 and 5 using equation (3). Panel A contains students' standardized follow-up one test scores, and Panel B contains students' demographic characteristics. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 4: Effects on In-School Reading

	Follow-Up One			Follow-Up Two		
	Control Mean	Treatment Effect No Controls	Treatment Effect Dist FE and Controls	Control Mean	Treatment Effect No Controls	Treatment Effect Dist FE and Controls
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Did student read any book (excluding	0.68	0.18***	0.19***	0.51	0.22***	0.22***
textbooks) in the last week in school?		(0.03)	(0.03)		(0.04)	(0.04)
No. of books (excluding textbooks) read in	1.85	2.35***	2.33***	1.34	0.98**	0.86
the last week in school		(0.52)	(0.56)		(0.49)	(0.53)
No. of books (excluding textbooks) read in	2.32	7.41***	7.17***	1.54	3.32***	3.12***
the last month in school		(0.88)	(0.82)		(0.56)	(0.42)
Is student able to give name of last book read	0.53	0.29***	0.30***	0.47	0.26***	0.26***
in school?		(0.04)	(0.03)		(0.04)	(0.03)
Is student able to describe last book read	0.55	0.22***	0.23***	0.46	0.23***	0.23***
in school?		(0.04)	(0.03)		(0.04)	(0.03)
Joint Test of All Differences						
$Chi^2(5)$		108.79	119.68		68.74	99.66
p-value		< 0.0001	< 0.0001		< 0.0001	< 0.0001

Note: This table presents a comparison of reading habits during the normal school day. The first three columns report results from the first follow-up survey, and the last three report results from the second. Columns 1 and 4 contain the control students' average responses. Columns 2 and 5 contain estimates of the treatment effect without any control variables. Columns 3 and 6 contain an estimate of the average differences, controlling for district fixed effects and demographic characteristics, namely age, gender, body mass index (BMI) (as well as their interactions), number of siblings, religion, language, and baseline reading score. The sample includes all students completing the respective follow-up tests, 5,228 for the first follow-up and 4,887 for the second. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 5: Effects on Reading Test Scores

Dependent Variable	Treatment Effect No Controls (1)	Treatment Effect Controls (2)	Treatment Effect Dist FE and Controls (3)
Follow-Up One	0.13*	0.12**	0.13***
	(0.07)	(0.05)	(0.05)
Follow-Up Two	0.08	0.07**	0.06**
	(0.07)	(0.03)	(0.03)

Note: This table displays the estimated effects of the treatment on students' reading test scores. Column 1 reports an estimate of the difference between the test scores of the control and treatment students. Column 2 reports an estimate of the average difference between the test scores of the control and treatment students controlling for demographic characteristics, namely age, gender, body mass index (BMI) (and their interactions), number of siblings, religion, language, and baseline reading score. Column 3 reports an estimate of the average difference between the test scores of the control and treatment students controlling for the aforementioned demographic characteristics and district fixed effects. The sample includes all students completing the respective follow-up tests, 5,228 for the first follow-up and 4,887 for the second. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 6: Effects on Reading Test Scores by Subject

	Follow-Up One		Follow-Up Two	
	Treatment Effect No Controls	Treatment Effect Dist FE and Controls	Treatment Effect No Controls	Treatment Effect Dist FE and Controls
Dependent Variable	(1)	(2)	(3)	(4)
Written Test	-0.02	-0.03	-0.05	-0.06*
	(0.05)	(0.04)	(0.05)	(0.03)
Letter Recognition	0.04	0.03	0.10*	0.07
	(0.05)	(0.05)	(0.06)	(0.04)
Sound Recognition	0.12	0.12	-0.02	-0.01
	(0.09)	(0.08)	(0.07)	(0.06)
Word Recognition	0.21**	0.19***	0.14**	0.12***
	(0.08)	(0.07)	(0.06)	(0.03)
Oral Reading	0.02	0.02	0.07	0.07*
	(0.05)	(0.03)	(0.06)	(0.04)
Oral Reading Questions	0.15	0.15*	0.07	0.06
	(0.10)	(0.08)	(0.07)	(0.05)
Average Score	0.13*	0.13***	0.08	0.06**
	(0.07)	(0.05)	(0.07)	(0.03)
Joint Test of All Differences				
$Chi^2(6)$	13.89	16.16	13.00	17.80
p-value	0.031	0.013	0.041	0.007

Note: This table displays the estimated effects of the treatment on students' standardized reading test scores. The first two columns report results from the first follow-up survey and the last two columns report results from the second follow-up survey. Columns 1 and 3 report an estimate of the average difference between the control and treatment students' test scores, without controls, from the first and second follow-up surveys, respectively. Columns 2 and 4 report an estimate of the average difference between the control and treatment students' test scores from the first and second follow-up surveys, respectively, controlling for district fixed effects and demographic characteristics, namely age, gender, body mass index (BMI) (and their interactions), number of siblings, religion, language, and baseline reading score. The sample includes all students completing the respective follow-up tests, 5,228 for the first follow-up and 4,887 for the second. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level. The last row of the table report the Chi^2 statistic and p-value from a test of the joint significance of all the coefficients on the treatment indicator from the regressions for the individual components of the reading tests estimated in a system of seemingly unrelated regression equations.

Table 7: Effects on Math and Social Studies Test Scores

Dependent Variable	Treatment Effect No Controls (1)	Treatment Effect Controls (2)	Treatment Effect Dist FE and Controls (3)
Math Score, Follow-Up One	0.04	0.04	0.03
	(0.07)	(0.05)	(0.05)
Social Studies, Follow-Up Two	0.06	0.05	0.06
	(0.06)	(0.05)	(0.06)

Note: This table displays the estimated effects of the treatment on students' math and social studies test scores. Column 1 reports an estimate of the difference between the test scores of the control and treatment students. Column 2 reports an estimate of the average difference between the test scores of the control and treatment students controlling for demographic characteristics, namely age, gender, body mass index (BMI) (and their interactions), number of siblings, religion, language, and baseline reading score. Column 3 reports an estimate of the average difference between the test scores of the control and treatment students controlling for the aforementioned demographic characteristics and district fixed effects. The sample includes all students completing the respective follow-up tests, 5,228 for the first follow-up and 4,887 for the second. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table 8: Effects on Reading Outside of School

	Follow-Up One			Follow-Up Two		
	Control Mean	Treatment Effect No Controls	Treatment Effect Dist FE and Controls	Control Mean	Treatment Effect No Controls	Treatment Effect Dist FE and Controls
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Did student read any book (excluding	0.48	< 0.01	0.01	0.45	0.01	0.01
textbooks) in the last week outside of school?		(0.04)	(0.04)		(0.04)	(0.03)
No. of books (excluding textbooks) read in	1.40	0.18	0.36	0.77	0.35***	0.34***
the last week outside school		(0.43)	(0.28)		(0.11)	(0.09)
No. of books (excluding textbooks) read in	1.61	1.28***	1.24***	1.20	0.92***	0.89***
the last month outside school		(0.30)	(0.25)		(0.18)	(0.16)
Is student able to give name of last book read	0.38	0.03	0.04	0.36	0.08**	0.07**
outside school?		(0.04)	(0.03)		(0.03)	(0.03)
Is student able to describe last book read	0.39	0.01	0.02	0.37	0.06*	0.06*
outside of school?		(0.04)	(0.03)		(0.04)	(0.03)
Joint Test of All Differences						
$Chi^2(5)$		36.67	38.26		71.19	75.18
p-value		<0.0001	<0.0001		<0.0001	<0.0001

Note: This table presents a comparison of reading habits outside of the normal school day. The first three columns report results from the first follow-up survey, and the last three report results from the second. Columns 1 and 4 contain the control students' average responses. Columns 2 and 5 contain estimates of the treatment effect without any control variables. Columns 3 and 6 contain an estimate of the average differences, controlling for district fixed effects and demographic characteristics, namely age, gender, body mass index (BMI) (as well as their interactions), number of siblings, religion, language, and baseline reading score. The sample includes all students completing the respective follow-up tests, 5,228 for the first follow-up and 4,887 for the second. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.

Table A1: Attrition Patterns, Follow-Up Two

	Non-Attritors			Non-Attritors Less Attritors		
Dependent Variable	Control Mean (1)	Treatment Mean (2)	Treatment Difference (3)	Control Difference (4)	Treatment Difference (5)	Difference-in-Difference (6)
Panel A: Standardized Baseline Reading Test Scores						
Written Test	0.02	0.03	0.01	0.17	0.05	-0.12
			(0.05)			(0.10)
Letter Recognition	0.01	0.05	0.04	0.10	0.16	0.07
			(0.05)			(0.07)
Sound Recognition	0.03	-0.08	-0.11	0.22	0.17	-0.05
			(0.07)			(0.09)
Word Recognition	0.03	0.07	0.04	0.27	0.44	0.17*
			(0.07)			(0.09)
Oral Reading	0.03	0.06	0.03	0.21	0.33	0.12
			(0.06)			(0.13)
Oral Reading Questions	0.01	0.06	0.05	0.08	0.18	0.11
			(0.07)			(0.10)
Average Score	0.03	0.04	0.01	0.24	0.31	0.07
			(0.07)			(0.10)
Panel B: Individual Characteristics						
Age	9.29	9.29	< 0.01	-0.57	-0.54	0.03
			(0.04)			(0.16)
Female	0.49	0.48	-0.01	0.10	0.10	0.00
			(0.01)			(0.04)
Height	128.18	128.23	0.05	-2.12	-1.49	0.62
			(0.32)			(0.73)
Weight	56.33	57.23	0.90	-1.89	-1.28	0.61
			(0.81)			(1.36)
BMI	15.42	15.65	0.23	0.01	0.04	0.03
			(0.18)			(0.23)
Siblings	3.83	3.91	0.08	-0.37	-0.36	0.02
			(0.11)			(0.17)
Catholic	0.74	0.69	-0.05*	0.04	0.02	-0.02
			(0.03)			(0.05)
INC	0.12	0.13	0.01	-0.04	-0.04	0.00
			(0.02)			(0.04)
Aglipayan	0.02	0.03	0.01	-0.01	0.00	0.01
			(0.01)			(0.01)
Born Again	0.07	0.08	0.02	0.01	-0.01	-0.02
			(0.01)			(0.02)
Protestant	0.03	0.04	0.01	0.00	0.01	0.01
			(0.01)			(0.02)
Other Religion	0.02	0.02	0.01	-0.01	0.01	0.02*
			(0.01)			(0.01)
Filipino	0.44	0.45	0.01	-0.02	0.01	0.03
			(0.07)			(0.06)
Iloco	0.19	0.15	-0.04	-0.01	0.04	0.05
			(0.05)			(0.04)
Kapampangan	0.37	0.40	0.03	0.03	-0.03	-0.06
			(0.09)			(0.08)
Pangasinan	0.01	< 0.01	< 0.01	< 0.01	-0.01	-0.01
			(< 0.01)			(0.01)
Other Language	< 0.01	< 0.01	< 0.01	< 0.01	-0.01	-0.01
			(< 0.01)			(0.01)
Observations	2280	2607	4887	2596	2914	5510

Note: This table presents a comparison of the attrition rates bewteen the treatment and control groups for the second follow-up survey. Column 1 contains the average characteristic of the baseline control students who took the second follow-up survey, while Column 2 contains the average characteristic of the baseline treatment students who took the second follow-up survey. Column 3 presents an estimate of the average difference in characteristics between the baseline control and treatment students who took the second follow-up survey. Columns 4 and 5 present the average difference in characteristics between the baseline students who took the second follow-up survey (non-attritors) and those who did not (attritors) for the control and treatment groups, respectively. Column 6 contains an estimate of the difference between the average differences in Columns 4 and 5. Panel A contains students' standardized follow-up one test scores and Panel B contains students' demographic characteristics. Standard errors are clustered by school and shown in parentheses. * indicates statistical significance at the 10% level, ** at the 5% percent level, and *** at the 1% level.