NBER WORKING PAPER SERIES

SCHOOL LIBRARIES AND LANGUAGE SKILLS IN INDIAN PRIMARY SCHOOLS: A RANDOMIZED EVALUATION OF THE AKSHARA LIBRARY PROGRAM

Evan Borkum
Fang He
Leigh L. Linden

Working Paper 18183
http://www.nber.org/papers/w18183

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2012

This project could not have been completed without the assistance of many individuals. We are indebted to Arvind Venkatadri, Dr. K. Vaijayanti, Ashok Kamath, Shreedevi Sharma, Pratima Bandekar, and many other staff members at the Akshara Foundation. Sadly, Pratima Bandekar, a colleague on many of our research projects, passed away during the course of this project. Her plucky, positive spirit is greatly missed, and we dedicate this effort to her. Finally, we gratefully acknowledge financial support from the Michael and Susan Dell Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

School Libraries and Language Skills in Indian Primary Schools: A Randomized Evaluation of the Akshara Library Program

Evan Borkum, Fang He, and Leigh L. Linden

NBER Working Paper No. 18183
June 2012
JEL No. I21,I28,O15

ABSTRACT

We conduct a randomized evaluation of a school library program on children's language skills. We find that the program had little impact on students' scores on a language test administered 16 months after implementation. The estimates are sufficiently precise to rule out effects larger than 0.13 and 0.11 standard deviations based on the 95 and 90 percent confidence intervals. The finding of zero effects is robust to different modes of implementation, individual tested language competencies and various subsets of the student population. We also find no impact on test scores in other subjects or on school attendance rates.

Evan Borkum
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
eborkum@mathematica-mpr.com

Fang He
General Accountability Office
441 G Street, NW
Room 4G48
Washington, DC 20548
hef@gao.gov

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

Section I: Introduction

The average quality of public schools in most developing countries is notoriously poor. This issue is receiving renewed attention given the increasing focus on improving the quality of education, which is explicitly listed as part of the second Millennium Development Goal of achieving universal primary education. The concern is that, while many countries are succeeding in increasing enrollment, the schools in which their children enroll are not suited to meet their educational needs (see, for example, Glewwe and Kremer, 2006). This problem is only compounded by the need to stretch already limited resources to serve the burgeoning demand. Identifying viable strategies for improving school quality is therefore an important policy issue.

We evaluate a promising educational program in Bangalore, India, that provides high quality libraries to public primary schools (our focus is on children in grades 3-5 in these schools). The intervention replaces existing libraries, which are typically poorly resourced, fail to contain age-appropriate reading material and are inaccessible to children, with libraries which are well-equipped with books designed to support the existing school curriculum and staffed with a dedicated librarian. In addition to regulating access to the collection, the librarian also provides regular reading-focused educational activities. These services are provided through a hub and spoke system that consists of a hub school that houses the library and a satellite spoke school without a library. A hub-based librarian then provides direct services to the hub school and transports reading material to the spoke schools.

Using a randomized controlled trial (RCT), we evaluate the program with a sample of 386 schools randomly selected from all of the public schools in Bangalore.

Schools are pre-assigned as either hubs or spokes based on geographic distribution and the availability of space for a library. The 200 hub schools and 186 spoke schools that we randomly selected from the universe of schools in the city were equally divided into respective treatment and control groups. Students' language skills were then evaluated using a baseline and a follow-up test administered 16 months later. We also tested students' math and science skills to check for ancillary benefits (or costs) to other subjects, collected attendance data to assess the effect of the program on participation, and had access to all administrative records on the students' use of the libraries.

Analysis of the library-use data suggests that the program was successfully implemented. Treatment was provided to all of the targeted schools[2] and within these schools an average of 81 percent of students utilized the libraries each month when schools were in session. This resulted in an overall participation rate for the average child of 2.41 visits a month and a borrowing rate of 1.26 books a month. Unfortunately, the intervention seems to have little effect on students' language skills. Overall, we find little average difference in the language scores of students in the treatment and control groups. In fact, the standard errors are tight enough that we can rule out effects larger than 0.13 and 0.11 standard deviations based on the 95 and 90 percent confidence intervals.

The finding of no effect is the same for all subsets of the data that we analyze. Neither hubs nor spokes show effects when analyzed individually. We also find no effect on individual tested language competencies (such as reading comprehension and vocabulary) and no effect on specific subsets of the students (i.e. by grade, baseline test score, demographic characteristics, etc.). Nor do we find differences in students' scores

[2] In fact there was a single school in our sample which refused a library. However, since we were also denied permission to test there, this was one of the schools removed from our sample as described in Section III.A.

on other subjects of the follow-up exam or in students' average attendance rates. These results suggest that supplemental programs like these may not be sufficient to significantly change students' test scores.

This study directly builds on an existing literature that investigates the education production function, specifically the effect of changes in the educational environment and resources on students' outcomes. Glewwe and Kremer (2006) provide a review of the literature in developing countries. The studies that they review generally find that the impact of school and teacher characteristics on student test scores is insignificant. Overall, the more convincing estimates from these evaluations imply that the success of any particular input-based educational intervention in developing countries largely depends on the characteristics of the program itself and its ability to operate effectively in spite of the problems inherent in the education systems of these countries. For example, Glewwe, Kremer and Moulin's (2007) study of textbooks and Glewwe et al.'s (2004) study of flip charts, both in Kenya, find these resources to have little overall impact on test scores. Programs that both provide resources and make a substantive change to the existing pedagogy, however, (such as Banerjee et al., 2007) do seem to have significant effects.

More specifically, a number of studies focus on the effects of programs directed at improving students' reading skills. In India, Banerjee et al. (2007) and He, Linden, and MacLeod (2009) find that low-cost remedial programs using young women from the community to teach basic literacy skills to young children substantially improved their

performance on a language test. Finally, Abeberese et al. (2010) find significant effects on students reading scores of a Philippine read-a-thon program.[3]

Our results raise doubts about the effectiveness of improving school libraries alone as viable educational interventions in the developing country context. The libraries in our study had little impact on academic achievement despite the fact that they made a significant change in the level and quality of available resources and involved pedagogical methods designed to support the existing curriculum. This is not to say that libraries cannot play an important role, but rather to suggest that (as in Glewwe and Kremer, 2006) to raise the possibility of doing so, libraries must do more than simply provide additional resources and bi-monthly programming.[4]

The remainder of the paper is set out as follows. Section II describes the library program in detail. Section III discusses the research design, including the data collection and the procedure for analyzing the data. Section IV checks that the study is representative of the wider population and is internally valid, while Section V estimates the impact of the library program on test scores and attendance. Section VI concludes.

Section II: The Akshara library program

The library program that we evaluate is run by the Akshara Foundation, a Bangalore-based NGO with the "mission to ensure that every child is in school and learning well". Because not all schools have sufficient space to host a library, Akshara libraries are

[3] Researchers in other fields have also contributed to this literature (i.e. Kim, 2007, Wassik and Slavin, 1993), including evaluations of various library programs in the United States, such as those in Alaska (Lance, Hamilton-Pennell and Rodney, 2000), Pennsylvania (Lance, Rodney and Hamilton-Pennell 2000b), New Mexico (Lance, Rodney and Hamilton-Pennell, 2002), North Carolina (Burgin and Bracy, 2003), Illinois (Lance, Rodney and Hamilton-Pennell 2005) and Texas (Smith, 2001).

[4] This is similar to the cited conclusions of the US literature on the effectiveness of libraries where programs which are more carefully integrated with existing curricula seem to be more effective.

organized according to a hub and spoke system with each hub school attached to several spokes in the same geographical area. Hub schools contain physical libraries that are set up in a room within the school and are staffed by a designated librarian. These librarians were recruited externally by Akshara and underwent several group training sessions after recruitment as well as several refresher sessions throughout the period of the study.

Training focused on library operations and on conducting activities in the libraries. The librarian's role in library operations includes maintaining the library, issuing books to students and keeping borrowing records. Books are color-coded by difficulty into six levels, and the librarian periodically evaluates children in order to decide whether they have sufficiently improved to progress to the next level (children select their own books within a level). The activities which the librarian was trained to implement included storytelling, role-playing games (where children act out a story from a book) and other educational games (such as identifying the sounds made by various animals in a story book). These activities are conducted during regular library periods that are held for each class according to a fixed timetable. The library periods also serve as the main opportunity for students to borrow books.

Spoke schools do not have a physical library. Instead, they are visited regularly by a mobile librarian according to a timetable.[5] The mobile librarian transports books from the hub library and issues them to the children using the same color-coding system as used as in the hub libraries. Typically, the mobile librarian will spend several hours at a spoke school, serving students in one class at a time. Unlike in the hub libraries, the

[5] The mobile librarian is often the hub librarian who visits the attached spoke schools when s/he is not engaged in duties at the hub. In some cases, particularly when the hub is very large so that the hub librarian does not have time to visit the spokes, a specialist mobile librarian is used.

librarians did not conduct activities in the spoke schools for the majority of the study period, although these activities were introduced in the final three months.

Data collected prior to the experiment (and described in Section III.B below) allow us to contrast the treatment with the resources that existed in control schools. Many schools did, in fact, already have some form of a library: approximately 94 percent of hub schools and 75 percent of spoke schools (85 percent of schools overall). However, the nature of these libraries differed dramatically from the Akshara libraries.

First, the new libraries were better resourced, better organized and more accessible than existing libraries. The new hub libraries were set up in a separate designated room whereas the old libraries usually had books located in a cupboard in a regular classroom or an office. The books supplied to the new libraries were also carefully selected to be appropriate for the targeted children and thus support the existing language curriculum whereas the quality and relevance of books in the old libraries was quite variable. In addition, the new libraries were organized around the color-coded difficulty system, making it easier for the children and librarians to select appropriate material and for the librarian to chart and monitor each child's progress through the difficulty levels. Finally, the new libraries were also accessible. Children were brought into the library on a regular schedule for activities and to borrow books.[6]

The second important improvement was the presence of a designated librarian in the new Akshara libraries: very few of the pre-existing libraries had such a librarian. For example, while overall 85 percent of total schools claimed to have some sort of library before the experiment, only 6 percent of total schools had a librarian. Without a dedicated

[6] The libraries in some schools were also available to children outside of the scheduled library period. Any non-scheduled visits are not included in our utilization data (discussed below), although any book borrowing during these visits is included.

librarian, access to resources would, at best, be provided to the entire school by a teacher simultaneously attending to the normal responsibilities of a teacher. This may have contributed to the fact that most of these resources were reputably inaccessible. Since the librarian plays a crucial role in mediating the interaction between students and the library[7] and also conducts educational activities in the hubs, one might have expected the new libraries to be more effective than the pre-existing ones in improving children's reading abilities.

As part of the program for monitoring the lending of books, we were able to obtain information on the total number of visits to each Akshara library in the study as well as the total number of books borrowed each month. These data confirm that students largely interacted with the libraries at the intended levels of intensity. Figures 1 and 2 provide a sense of the degree of interaction between children and the program libraries over the period of the study. For each month, Figure 1 illustrates the average fraction of children in our sample of schools that visited the library at least once during library period and the fraction that borrowed at least one book.[8] The average fraction of children visiting the library in each month of the study fluctuates between 0.5 and 0.9 in most months. The exceptions are February 2008 (school exams) and May 2008 (summer vacation), when virtually no children visited the libraries. The fact that not all children visit in each month could be the result of several factors. It could be that some children are absent on the days on which library visits are scheduled or that certain library periods

[7] As discussed earlier, Lance, Rodney and Hamilton-Pennell's (2000a) study of Colorado public schools finds that the librarian plays an important role in determining the impact of libraries on test scores. Beyond the librarian's role as an instructor, collaboration between the librarian and the rest of the school staff and management is identified as a key feature of successful libraries.

[8] The data in Figures 1 and 2 are only available in aggregate for children in grades 2-7 (with slight variation based on the grade span of the school), while the evaluation targets children in grades 3-5. However, one would expect a similar pattern for children in the targeted grades since there were no obvious differences in program implementation across grades.

in a month are canceled (e.g. because of a religious holiday or because the librarian was absent).

Figure 1 also shows that the fraction of children borrowing at least one book in a month begins at a very low level but increases by the end of 2007 as the stock of books arrived in the libraries. This fraction again fluctuates across months, with notable drops in February 2007 and May 2007 as before. In general the fraction borrowing is smaller than the fraction visiting, suggesting that not all students who visit the library during library period are borrowing books. However, in most months over 60 percent of students borrow at least one book a month.

Figure 2 illustrates the average number of library period visits and average number of books borrowed per child in each month. These monthly averages are unconditional means for all children in grades 2-7. Not including the initial months when the libraries were starting or the months with significant holidays or exams (Sept/Oct 2007, Feb 2008, and May 2008), children in the average school visit a mean of 2.41 times a month and borrow a mean of 1.26 books per month. These mean monthly usage figures are slightly higher in the hubs, with an average of 2.65 visits and 1.35 books per month compared to 2.12 visits and 1.16 books per month in the spokes.[9,10]

In addition to the evidence from Figure 1, data on the average number of visits conditional on at least one visit a month (not shown) confirm that overall participation is

[9] Data on utilization rates by implementation method available upon request.

[10] There is some variation in usage rates across schools: the mean number of visits per month has a standard deviation of 1.32 and the mean number of books borrowed per month has a standard deviation of 0.87. We attempted to estimate whether the treatment effect varies by usage rates by constructing predicted usage rates in all schools using baseline school characteristics that are correlated with usage in the treatment schools. However, we did not find any significant variation in the treatment effect by this predicted usage measure.

quite high. The conditional mean for the number of visits is 3.06 per month, which is only 0.65 visits per month higher than the unconditional mean.

Section III: Research design

A. Research groups

To evaluate the impact of the library program on students' academic achievement, we implemented a randomized controlled trial. All government primary schools in Bangalore were initially arranged in a hub and spoke pattern by Akshara. Hubs were chosen based on size, geographic location and the availability of a room to house a library. The remaining schools were attached as spokes to a nearby hub, with each hub attached to up to seven spoke schools. While hub schools usually have fairly high enrollment rates, spoke schools have both lower average enrollment levels and significantly more variation in size. Because of this significant variation, we chose to trim the sample both to eliminate outliers and to remove schools that had far too few students to include in the study. We chose to exclude the upper and lower 10 percent of spoke schools, removing those with more than 230 students and less than 20 students. In order to implement our randomization strategy, we also trimmed the smallest 5 hub schools (those with less than 17 students) so that we were left with a population of exactly 300 hubs in the city.

We then randomly selected a single spoke for each hub, obtaining a sample of 300 hub-spoke pairs that we refer to as "units".[11] Given the available budget and power calculations conducted using reading tests available from the previous year, we

[11] The elimination of the smallest 10 percent of spoke schools meant that a small number of hubs (those attached only to small spokes) were not assigned a spoke.

determined that only 200 units would be included in the study.[12] As a result, we designed the randomization both to select a random sample for the study and to assign units either to a treatment or a control group. Specifically we followed a "matched-pair" randomization strategy in which units were first grouped by geographic location (we used "blocks" which are most analogous to US zip codes), ranked by the average test score of the unit within each location, and then grouped into triplets.[13] Within each triplet, one unit was selected for the treatment group, one for the control group, and one to be left out of the study with equal probability.[14]

Immediately after the randomization but before the initiation of the treatment, we began conducting the baseline test in the 200 units selected for the study. However, we were unable to conduct the baseline test in 16 schools forcing us to remove them from the sample.[15] As a result, the final sample included 193 units comprising 193 hubs and 177 spokes. Within each school, we randomly chose one class per grade to participate in the study in the case that a grade contained more than a single class.[16] The study includes students in grades 3 through 5: grades were only omitted if the included school did not offer a particular grade.

Table 1 describes the resulting sample of schools and students. Panel A shows the distribution by school, class, and student while Panels B and C show the distribution by

[12] We chose the sample to be large enough to yield a minimally detectable effect size of 0.15 standard deviations with 90 percent power and a 5 percent significance level, using a bootstrap procedure based on data (including language scores) from the previous academic year.
[13] Unmatched units were then grouped together and matched by average test score irrespective of geographic location.
[14] We did not employ a strategy that utilized re-randomization if the resulting research groups were unbalanced.
[15] In results available upon request, we demonstrate the balance by research group of the participating schools using the variables available on all city schools in Table 2.
[16] As explained in Section III.C., we re-weight the schools by the number of classes to ensure that our random sample is representative of all students in the city.

grade and gender. Starting with Panel A, the research groups have similar numbers of schools, classes, and students. The medium of instruction in schools in our sample is either Kannada or Urdu: column 3 of Table 1 shows that 90% of the schools were Kannada medium. Panel B shows that students are also equally distributed by grade and Panel C shows a similar distribution by gender. In total, our sample includes 20,858 students – 14,455 in hub schools and 6,403 in spoke schools.

B. Data

Three main sets of data were collected for the evaluation. First, we had access to data on all of the schools in the city that had been collected by Akshara prior to this study. This included information on the resources available in every school including libraries and librarians as well as other amenities such as access to water and availability of toilets. As mentioned in the description of the randomization, these data also included a reading test that was conducted at the end of the previous academic year in all schools.

Second, we administered our own baseline test and a follow-up test 16 months later. We conducted a baseline assessment at the start of the 2007-2008 academic year (July-August 2007) prior to the start of the libraries. The baseline assessment consisted of a written test assessing students' language skills and basic math skills which was administered in the medium of instruction of the school (Kannada or Urdu). In December 2008 (a few months before the end of the 2008-2009 academic year[17]), we conducted follow-up tests in the treatment and control schools. These tests were specific to each grade, and included sections on language, math and environmental science (EVS).

[17] We would have preferred to test at the end of the academic year, but we were required to complete all testing before the schools began administering the end of year exams.

The baseline and follow-up tests differed in both the subjects that were covered and the comprehensiveness with which the included subjects were covered. The baseline test was meant to be a quick diagnostic tool that could be used to check the comparability of the research groups at the start of the study and to control for student ability at baseline when comparing the average post-test scores. It included a few basic language and math competencies such as letter, word, and picture identification, sentence completion, story comprehension, counting, identifying number patterns, simple arithmetic and word problems. The follow-up test was designed to constitute a much more comprehensive assessment of student performance in the three main subjects offered by the Bangalore public schools. We recruited a team of senior teachers who compiled a list of all competencies covered by the official state of Karnataka curriculum for grades one through six along with sample questions. We then piloted the questions and created a separate exam for students in grades 4, 5 and 6 (our initial grade 3, 4 and 5 samples had advanced by a year) by eliminating the questions that yielded little variation in student performance.[18]

The main outcome of interest is a student's score on the language section of the follow-up test. While the Akshara libraries might have been expected to primarily influence students' reading skills, we chose to include a wider range of language competencies rather than reading skills alone (such as grammar, punctuation and vocabulary) because reading is likely to influence performance in these other competencies. In addition, since language skills are generally applicable, Math and EVS

[18] Because many children perform below their grade level, we could not simply test all of the competencies in the respective grade. In addition, because the purpose of the experiment is to compare the performance of children in the treatment and control groups (rather than measuring the overall level of achievement), including questions that all children answered either correctly or incorrectly would add little value to the exam.

were included to assess the possibility that student with better language skills were also better equipped to learn other subjects. This also allowed us to assess the possibility that the focus of the program on language skills may have come at the expense of time devoted to other subjects.

Finally, we collected two types of administrative information from the schools. The first consisted of basic demographic information on the students in the sample, which is available from the school admissions registers. This includes information on gender, age, mother tongue, religion and caste. The demographic information is useful both for ensuring comparability between treatment and control groups and as controls in the follow-up comparisons to improve precision. It also allows one to break down the results by certain demographic categories in order to examine possible heterogeneous effects (by gender, for example). The second type of administrative information consisted of monthly attendance records (number of days attended), collected from daily class attendance rosters by our surveyors. Since the improvement in the quality of the libraries may have encouraged school attendance, we are interested in potential effects along this dimension. Attendance information was only collected for the first 7 months of the study, from September 2007 to March 2008. Such data must always be used with caution since teachers may incorrectly record attendance data when they have an incentive to do so (Shastry and Linden, 2009). However, in this context, the intervention does not change teachers' incentives to report students as being either absent or present.

C. Analytic models

The basic research design involves a comparison of students in the treatment and control groups. The advantage of an RCT is that, by construction, the assignment of schools to the treatment is independent of school and student characteristics. This statistical independence ensures that the primary difference between the two groups is their receipt of the treatment rather than other factors that could be correlated with the students' follow-up test scores.

If the treatment assignment is statistically independent of student and school characteristics, then we can directly estimate the effects of the intervention by comparing the average outcomes in the treatment and control groups using the following equation:

$$Y_{ijk} = \beta_0 + \beta_1 Treat_k + \beta_2 X_{ijk} + \varepsilon_{ijk} \quad (1)$$

where Y_{ijk} is the outcome for student i in grade j in school k (such as gender, mother tongue or baseline test score) and $Treat_k$ is an indicator that is one if school k is treated and zero if it is not.[19] This specification also includes the vector of control variables X_{ijk}. The controls used here include baseline test scores (linear and quadratic terms), reading score of the hub-spoke unit (used in the stratification) and indicators for gender, grade, age, majority religion and language group affiliations, caste status and for having a mother tongue different from the medium of instruction of one's school.[20]

Because this is a clustered RCT, it is important to account for correlation of students test scores at the level of treatment assignment. Not doing so would cause us to

---

We normalize all test scores within grade and medium of instruction relative to the control group distribution for each exam. A student's normalized score therefore reflects her performance (in standard deviations), relative to other students in her grade in schools with the same medium of instruction.

[20] Many of these controls could not be obtained for all the students. In order to avoid dropping observations in regressions with controls, an additional category for each variable was created signifying a missing value.

over-estimate the precision with which the treatment effect can be estimated and possibly cause us to over-reject the null hypothesis of no effect. To account for this, the standard errors in all regressions are clustered at the unit level for regressions that include all schools and the school level when the hub and spoke samples are used individually.

Section IV: Internal and External Validity

We aim to estimate the causal effect of the Akshara library program on schools in Bangalore. To do so, the research design must satisfy three criteria. First, the schools in our sample must be a representative subset of schools in the city so that the results are externally valid in the sense that they can be generalized to the full set of schools. Second, the randomization must succeed in assigning schools to the treatment group independently of student and school characteristics. Finally, the follow-up survey must have been administered in such a way that the ability to observe students is also independent of the treatment assignment.

A. External Validity

The randomization was designed both to assign schools to the research groups and to select a representative subset of schools from all of the schools in Bangalore. In order for our sample to be representative, the inclusion of units into the study must be statistically independent of school and student characteristics. To check this, we use the school characteristics from the data obtained by Akshara prior to the study to compare schools included and not included in the study.

Table 2 contains the results of these comparisons. For each of the characteristics we provide the mean of schools in the sample, the mean for out of sample schools and finally the difference between the two. The first three columns contain the estimates for all of the schools. The second three contain estimates for only hub schools, and finally, the last three columns contain estimates for the spoke schools.

Overall, the estimates are consistent with the independence of selection into the sample. Only one of the differences is statistically significant – the difference in the probability that a library exists before the treatment – and even this is only a difference of around 6 percentage points for all schools and is only marginally significant. Looking across the remaining columns, it is clear that this difference is driven primarily by the spoke schools. The difference for the hubs is almost exactly zero while the difference for the spoke schools is about 13 percentage points.[21] The differences observed in all other characteristics are both small and statistically insignificant for all schools, hub schools, and spoke schools.

B. Internal Validity: Baseline Comparison

If the assignment of the treatment was statistically independent of the characteristics of the schools and the students therein, then any differences between treatment and control groups should be due to random variation alone. We can check that this is the case by comparing the two groups using the characteristics observed in our data set. As expected,

[21] While this is the only significant difference, it is consistent with a larger pattern in the data which is that the random assignment process generated more comparable research groups for the hub schools than the spoke schools. None of the differences are large enough to threaten the validity of the study, but it most likely results from the way in which the units were stratified. Spoke schools are generally much smaller than the hub to which they are attached. As a result, the average reading score of the unit used in that stratification was more representative of the hub than the spoke. This probably resulted in the randomization imposing a greater constraint on the similarity of the hubs than of the spokes.

the results are consistent with the schools having been allocated to treatment independently of school and student characteristics. Only a small number of the differences are statistically significant. The magnitudes of these differences are small enough that, when their correlation with the post test score is taken into account, they provide no reason to expect large differences in post-test scores unrelated to the provision of the treatment.

These estimates are performed using equation (1) without control variables and are presented in Table 3. As in Table 2, we present the results for all schools, hub schools, and then spoke schools in subsequent groups of three columns. For each group of schools, the first column presents the average characteristics for the control group and the second presents the average difference between the treatment and control groups. To assist in gauging the implications of the differences for the follow-up comparisons, the third column presents coefficients estimated by regressing the follow-up language score on each of the presented characteristics. Table 3 conducts these comparisons for baseline characteristics which include school level features (Panel A), baseline test scores (Panel B) and student demographics (Panel C).

The comparisons in panel A of Table 3 suggest that treatment and control schools contained a similar number of students in the targeted grades. The fraction with a pre-existing library and librarian are also not significantly different, nor are the fractions with running water and toilet facilities. Treatment and control schools were therefore similar along these dimensions, which include measures of the level of library resources at baseline.

Panel B compares the normalized baseline test scores of treatment and control students. For the overall sample in column 2 and for the hub schools in column 5, the magnitude of the differences in baseline test scores are small (0.09 standard deviations or lower) and statistically indistinguishable from zero. For the spokes, on the other hand, the magnitude of the differences in column 8 ranges from 0.24 to 0.28 standard deviations. These differences are all significantly different from zero, suggesting that the students in the control spokes performed better than those in the treatment spokes on the baseline test. However, given the correlation between baseline test scores and follow-up score in column 9, these differences are only equivalent to a 0.06-0.08 standard deviation difference in the follow-up score. Nevertheless, we will of course include baseline scores when estimating the difference in follow-up scores so that these small differences should not drive the results for the spokes.

As a final set of baseline comparisons, Panel C of Table 3 compares treatment and control students along a number of demographic characteristics that one might expect to be relevant for academic performance. These characteristics include gender, age, and indicators for being a member of the majority religion and language groups, being a member of a disadvantaged caste and having one's mother tongue different from a school's medium of instruction. Along all these dimensions, differences between students in treatment and control schools are small and statistically insignificant for the full sample and for hubs and spokes separately. Although this implies that the research groups are comparable in terms of student demographics, these characteristics will still be controlled for in the follow-up comparisons in order to improve precision since some of them (e.g. gender) are significantly correlated with follow-up scores.

C. Internal Validity: Attrition

Despite the imposed independence, a major threat to the validity of the design is the possible correlation between receipt of the treatment and the probability that a child is observable during the follow-up survey. Any correlation raises the possibility of selection by unobserved characteristics biasing the treatment estimates even after controlling for all observed characteristics. While it is obviously impossible to investigate selection by characteristics not observed in the data, it is unlikely that such selection would occur without also occurring by observed characteristics.

Tables 4 and 5 therefore estimate the attrition patterns based on observed characteristics. Table 4 contains a comparison of the students who took the baseline test but failed to take the follow-up test. Panel A contains the overall estimated probability that a child who took the baseline attrited from the sample. Overall the rate is high, with about 27 percent of all students in the baseline failing to take the follow-up exam. This is not unusual for school-based studies that track children across academic years. In addition, the rates within each group of schools are similar for treatment and control.[22] The largest difference is the statistically insignificant 3.9 percentage points difference for the spoke schools.

Panels B, C, and D of Table 4 are structured identically to Panels A, B, and C of Table 3. Consistent with the fact that students' presence during the testing session was independent of the treatment assignment, none of the differences in characteristics of the attriting students is likely to substantially bias the treatment estimates. Only the differences in baseline scores for the spokes (column 6) stand out as consistently

[22] In fact, some schools terminated at grade 6 so that fifth graders in these schools had to mechanically ascend to another school in the second year of the study and therefore dropped out of our sample.

statistically significant. However, these reflect the pattern observed at baseline rather than differential attrition and, given the correlations in Table 3, are unlikely to threaten the validity of the study.

Table 5 presents the same analysis as in Table 3 but using only those students who took the follow-up test. This presents the net effect of the attrition process on the sample and also provides a baseline comparison that includes only those students that will be used for the comparison of the follow-up tests. The results are almost identical to those presented in Table 3, confirming that the attrition process did not significantly change the characteristics of the sample, and providing support for attributing the difference in test scores at follow-up to the library intervention.

Section V: Treatment Effects

This section estimates the impact of the Akshara library program by comparing the outcomes of interest in treatment and control schools. First, we compare follow-up test scores. We are interested primarily in students' scores on the follow-up language section of the exam, but we are also interested in the scores on other sections of the follow-up exam (math and EVS) in case there were spillovers into these other subject areas. Second, we check whether the estimated impact on test scores for the full sample is masking differential effects among certain subsets of students. Finally, we investigate whether the library program had an impact on the rate at which students attend school.

A. Impact on Test Scores

Our primary outcome of interest is students' score on the language section of the exam. Table 6 presents these results for the follow-up test scores, reporting the effects for the scores in each individual section of the follow-up test (language, math and EVS) and then the aggregate score. Estimates are provided for the entire sample and then just for the hubs and spokes as in previous tables. For each group, the first column provides the simple difference between the treatment and control group using equation (1) without the control variables while the second columns provides the difference estimate with the controls.

Overall, the program seems to have had little impact on students' language scores. Focusing on the entire sample, the magnitude of the difference in normalized test scores for this section is only 0.002 standard deviations using the simple difference and 0.004 standard deviations controlling for baseline characteristics. The overall estimate of the program is also fairly precise. The standard error on the estimate with controls in column 2 is 0.070 standard deviations, yielding a 95 percent confidence interval of (-0.14, 0.13) and a 90 percent confidence interval of (-0.12, 0.11). The estimated difference with controls for the hub sample in Table 6 is similarly small (0.028 standard deviations), while the estimate for the spokes is negative in sign (-0.12 standard deviations) but not statistically distinguishable from zero.

Recently, many viable interventions have yielded effects in the range of 0.15 to 0.5 standard deviations. For example, the computer-assisted learning program in Banerjee et al (2007) increased test scores by 0.21 standard deviations, the English-language program in He, Linden, and MacLeod (2008) yielded gains of 0.25-0.35 standard

deviations, and a community preschool intervention increased reading scores by between 0.12 and 0.70 standard deviations (He, Linden, and MacLeod, 2009). We can reject the hypothesis that the treatment effect of the program is in the same range of these successful programs.

The estimated effect on the other subjects and the total score are similarly small. For each variable, the estimated differences are all in a similar range to the differences in language scores and none of them are statistically significant. The lack of a positive effect suggests that the program did not have positive spillovers to other subjects – not surprising given the lack of a direct effect on language skills. However, the lack of a strong negative effect also indicates that any time spent in the library did not come at the expense of the other subjects.

While the program may not have had an effect on test scores overall, it is possible that it may have impacted certain competencies. To test this, we break the language test up into the four major competency groups from the state curriculum. Table 7 lists these results. The format for this table is identical to that of Table 6.

The results are remarkably consistent with the overall averages presented in Table 6. With the exception of grammar and reading comprehension for the spoke schools, all of the differences controlling for baseline characteristics are less than a tenth of a standard deviation in magnitude. The differences for both of these competencies are statistically significant in the simple difference for the spoke schools – and actually negative in sign – but such variation should be expected when subdividing the results so finely. These results are consistent with the overall test scores in table 6 and suggest that

the library program is not effective at improving student's acquisition of the language competencies specified by the official curriculum.

B. Impact on Test Scores: Heterogeneous Effects

The overall result of no impact on test scores could be masking effects on particular subgroups of the student population. Some groups of students may have experienced positive or negative effects that are diluted or offset by the effects experienced by other groups. In order to investigate whether this is the case, we re-estimate the impact of the program on the follow-up language test score by dividing our sample of students along several dimensions in Panels B-F of Table 8. These dimensions include gender, grade, performance on the baseline language test (as a quartile of the overall distribution), demographic characteristics and library resources in the school at baseline. All these dimensions are salient in that students with these characteristics could plausibly have different responses to the library program. For each group of schools (all schools, hubs and spokes), the difference in language test scores is presented in the second column and includes controls for baseline characteristics.

The results in column 2 of Table 8 fail to find any significant impact of the program on test scores for any subgroup that we consider in the overall sample. The estimates in this column are all small and statistically insignificant. Focusing on the hub school students in column 4, the impacts for each subgroup in panels B-E are similarly small and insignificant. For the spoke students, some of the point estimates in column 6 are somewhat larger in magnitude (up to 0.19 standard deviations) but are quite imprecise and are all negative in sign.

While existing libraries are rarely used, one might still expect the largest effect of the program in schools without existing libraries since the change due to the program is likely to largest in these schools. However, the results presented in Panel F suggest that the program is not effective in these schools either. The overall treatment effect for these schools is -0.022 standard deviations. The treatment effect for hub schools alone is large and positive at 0.38 standard deviations, but statistically insignificant. Conversely, the effect for spoke schools without a library is large and negative, but statistically insignificant. As a result, the evidence in Table 8 provides little support for the hypothesis that the library program had a positive impact on language scores for any of the subgroups that we consider.

C. Impact on attendance

It is possible that the presence of a new library affected student attendance rates – for example, the library may have served as an incentive for students in treatment schools to attend school more regularly. In addition, while we fail to observe differences in test scores, attendance records were obtained to better understand the mechanism by which the program might have improved students' performances. If, for example, treatment children attended at a higher rate, then the fact that these students were simply in school more often could have also contributed to higher test scores.

In order to investigate whether the library had an impact on attendance, we run estimates of the form of equation (1) with the attendance rate as the dependant variable and without other controls. This variable is just the number of days a student was marked

as present in the school register over the entire study period as a fraction of the number of days on which the school was open (i.e. excluding weekends and holidays).

The results are presented in Table 9, with the results for the full sample in Panel A and the results for various sub-groups in Panels B-F. The overall attendance rates are very high: the means for the controls schools suggest that average attendance rates in these schools were around 90%, with little variation in this mean among the various subgroups. The coefficients in columns 2, 4 and 6 are all close to zero and statistically insignificant, suggesting that the library program had little impact on attendance rates. This is consistent with the results from other interventions in India, in which improvements in the quality of the learning environment do not seem to improve attendance (Banerjee et al., 2007; He, Linden and MacLeod, 2008).

Section VI: Conclusion

This study provides an evaluation of the effects of a library program for primary school children in India after 16 months of implementation. The program provides both high quality reading material designed to support the existing language curriculum and direct librarian activities using a pedagogically distinct strategy from what students normally experience. The program provides services to selected schools in the city through a hub-and-spoke system in which some schools are served directly and other schools are served indirectly from the schools receiving direct treatment. We find that, both overall and for each mode of treatment implementation, the program had no effect on students' language skills or on their performance in other subjects. The results are remarkably consistent across individual language competencies (grammar, reading comprehension, punctuation,

and vocabulary), mode of implementation, and also within individual subsets of the student population (gender, grade, baseline test score, demographic characteristics, and existing school resources). Overall we can reject treatment estimates of 0.13 and above at the 95 percent confidence level and of 0.11 and above at the 90 percent level.

The consistency of the results suggests a problem with the treatment itself rather than a mismatch between the program and the needs of particular students and schools. The results also stand in contrast to many recent studies in developing countries that showed large positive effects of programs that provided additional resources while teaching students using a pedagogical methods that is different than they normally experience. The main difference between this study and those showing a significant effect is the intensity of the treatment. Students in this study interacted with the librarian only twice a month on average while students in the other studies interacted with the new teaching methods several times a week. We know from a similar set of studies that resources alone usually are not enough to move test scores (Glewwe and Kremer, 2006), suggesting that to be effective the librarians may need to have significantly more contact with students. This also suggests that an important area for future study is to understand the relationship between the intensity of student interaction and the effects of an intervention.

Section VII: References

Abeberese, A., T. Kumler, and L. Linden (2010) "Improving Reading Skills by Encouraging Children to Read: A Randomized Evaluation of the Sa Aklat Sisikat Program in the Philippines," mimeo: The University of Texas at Austin.

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.

Burgin, R. and Bracy, P.B. (2003); "An Essential Connection: How Quality School Library Media Programs Improve Student Achievement in North Carolina"; RB Software & Consulting.

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

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.

He, F., Linden, L. and MacLeod, M. (2009), "Teaching Pre-Schoolers to Read: A Randomized Evaluation of the Pratham Shishuvachan Program"; mimeo: 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.

Lance, K.C., Welborn, L. and Hamilton-Pennell, C. (1992); "The Impact of School Library Media Centers on Academic Achievement"; State Library and Adult Education Office; Colorado Department of Education.

Lance, K.C., Hamilton-Pennell, C. and Rodney, M.J. (2000); "Information Empowered: The School Librarian as an Agent of Academic Achievement in Alaska Schools"; Alaska State Library.

Lance, K.C., Rodney, M.J. and Hamilton-Pennell, C. (2000a); "How School Librarians Help Kids Achieve Standards: The Second Colorado Study"; Colorado State Library; Colorado Department of Education.

--- (2000b) "Measuring Up to Standards: The Impact of School Library Programs and Information Literacy in Pennsylvania Schools"; Office of Commonwealth Libraries; Pennsylvania Department of Education.

--- (2002) "How School Libraries Improve Outcomes for Children: The New Mexico Study"; New Mexico State Library.

--- (2005) "Powerful Libraries Make Powerful Learners: The Illinois Study"; Illinois School Media Library Association.

Machin, S. and McNally, S. (2004); "The Literacy Hour"; IZA Discussion Paper No. 1005.

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.

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; 1995.

Shastry, G. and Linden, L. (2009) "Identifying Agent Discretion: Exaggerating Student Attendance in Response to a Conditional School Nutrition Program"; mimeo: Columbia University.

Smith, E.G. (2001); "Texas School Libraries: Standards, Resources, Services and Students' Performance"; EGS Research and Consulting.

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

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.

Table 1: Sample composition

	All			Hubs			Spokes		
	Control	Treatment	Total	Control	Treatment	Total	Control	Treatment	Total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: School, classes, students									
Number schools	188	182	370	96	97	193	92	85	177
Kannada medium	171	162	333	94	89	183	77	73	150
Number classes	553	541	1094	283	290	573	270	251	521
Number students	10960	9898	20858	7602	6853	14455	3358	3045	6403
Panel B: Grades									
Grade 3	3635	3215	6850	2480	2166	4646	1155	1049	2204
Grade 4	3675	3150	6825	2505	2182	4687	1170	968	2138
Grade 5	3650	3533	7183	2617	2505	5122	1033	1028	2061
Panel C: Gender									
Males	5350	4907	10257	3770	3422	7192	1580	1485	3065
Females	5610	4991	10601	3832	3431	7263	1778	1560	3338

Notes: Table shows composition of sample for full sample (columns 1-3), hubs (columns 4-6) and spokes (columns 7-9). Table omits 16 schools that could not be tested in the baseline test.

Table 2: External Validity

	All			Hubs			Spokes		
	Sample mean	Out-sample mean	Diff.	Sample mean	Out-sample mean	Diff.	Sample mean	Out-sample mean	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Library exists	0.849	0.907	-0.058*	0.94	0.93	0.01	0.751	0.883	-0.132***
			(0.030)			(0.030)			(0.051)
Librarian exists	0.062	0.046	0.017	0.076	0.054	0.022	0.044	0.037	0.007
			(0.022)			(0.032)			(0.028)
Computer exists	0.241	0.246	-0.005	0.333	0.303	0.03	0.139	0.185	-0.046
			(0.038)			(0.058)			(0.046)
Water	0.878	0.865	0.013	0.91	0.902	0.008	0.844	0.826	0.019
			(0.031)			(0.037)			(0.049)
Toilet	0.846	0.84	0.006	0.875	0.9	-0.025	0.815	0.777	0.039
			(0.032)			(0.039)			(0.051)
Kannada	0.899	0.86	0.039	0.945	0.929	0.016	0.849	0.787	0.062
			(0.028)			(0.029)			(0.048)
School reading score	0.055	0.067	-0.012	0.064	0.07	-0.006	0.045	0.064	-0.02
			(0.043)			(0.059)			(0.063)
N	386	195		200	100		186	95	

Notes: Table compares the baseline characteristics of the 200 hub-spoke units that were selected for the study (sample schools) and the 100 hub-spoke units (out-sample schools) that were not selected. Columns 1-3 are for all schools, columns 4-6 are for the hubs, and columns 7-9 are for the spokes. Columns 1, 4, and 7 list the means for the sample schools that were selected for the study. Columns 2, 5, and 8 list the means for the out-of-sample schools. Columns 3, 6, and 9 show the difference in means between the out-of-sample and sample means.

Table 3: Baseline Comparisons

	All			Hubs			Spokes		
	C mean	T-C	Posttest corr.	C mean	T-C	Posttest corr.	C mean	T-C	Posttest corr.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: School characteristics									
Number students	94.023	-6.622	0	110.203	-7.125	0.001	52.432	-4.678	-0.001
		(7.320)	(0.001)		(8.602)	(0.001)		(5.729)	(0.003)
Library exists	0.931	-0.037	-0.252*	0.979	-0.039	-0.051	0.806	-0.028	-0.264*
		(0.030)	(0.133)		(0.027)	(0.255)		(0.072)	(0.146)
Librarian exists	0.075	-0.035	0.139	0.078	-0.044	0.15	0.063	-0.007	0.08
		(0.031)	(0.159)		(0.036)	(0.186)		(0.053)	(0.307)
Water	0.877	0.022	-0.06	0.901	-0.003	-0.123	0.81	0.089	0.135
		(0.052)	(0.110)		(0.062)	(0.113)		(0.066)	(0.222)
Toilet	0.891	-0.05	-0.164*	0.928	-0.109*	-0.226**	0.793	0.104*	0.023
		(0.046)	(0.086)		(0.058)	(0.102)		(0.062)	(0.141)
School reading score	0.026	0.041	0.174**	0.046	0.022	0.119	-0.024	0.089	0.306**
		(0.058)	(0.083)		(0.067)	(0.098)		(0.087)	(0.137)
Panel B: Baseline test									
Language	0	-0.022	0.326***	-0.072	0.071	0.334***	0.185	-0.261***	0.291***
		(0.067)	(0.024)		(0.082)	(0.028)		(0.082)	(0.040)
Math	0	-0.009	0.281***	-0.079	0.08	0.283***	0.202	-0.239***	0.263***
		(0.068)	(0.024)		(0.081)	(0.029)		(0.088)	(0.041)
Total	0	-0.015	0.327***	-0.085	0.087	0.331***	0.219	-0.275***	0.301***
		(0.070)	(0.025)		(0.083)	(0.031)		(0.089)	(0.041)
Panel C: Individual characteristics									
Male	0.484	0.012	-0.188***	0.49	0.01	-0.193***	0.471	0.017	-0.168***
		(0.011)	(0.021)		(0.015)	(0.026)		(0.016)	(0.043)
Hindu	0.845	0.017	0.064	0.888	-0.005	0.075	0.739	0.069	0.134
		(0.034)	(0.077)		(0.035)	(0.072)		(0.073)	(0.151)
Kannada mothertongue	0.669	-0.016	0.123**	0.716	-0.052	0.148***	0.554	0.07	0.11
		(0.043)	(0.049)		(0.046)	(0.048)		(0.072)	(0.106)
Mothertongue/ medium different	0.238	0.022	-0.139***	0.252	0.033	-0.168***	0.203	-0.005	0.014
		(0.031)	(0.047)		(0.039)	(0.048)		(0.047)	(0.116)
Scheduled Tribe/Caste	0.565	0.022	-0.093	0.528	0.018	-0.197***	0.69	0.008	0.180*
		(0.047)	(0.057)		(0.054)	(0.058)		(0.064)	(0.102)
Age	8.9	-0.007	-0.006	8.885	0.025	0.022	8.936	-0.087	-0.075**
		(0.040)	(0.015)		(0.046)	(0.017)		(0.060)	(0.030)

Notes: Table compares baseline characteristics for all students who were tested at baseline. Comparisons are for the full sample (columns 1-3), hubs (columns 4-6) and spokes (columns 7-9). Columns 1, 4 and 7 show the mean of each variable for the control students. Columns 2, 5 and 8 show the difference in the mean of each variable for treatment and control students. Columns 3, 6 and 9 show the simple correlation coefficient between each characteristic and the follow-up language test score. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 4: Comparison of Attriting Students

	All		Hubs		Spokes	
	C mean	T-C	C mean	T-C	C mean	T-C
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Overall						
	0.272	0.01	0.232	0.029	0.374	-0.039
		(0.018)		(0.020)		(0.032)
Panel B: School characteristics						
Number students	87.467	-5.712	112.272	-10.993	47.949	-4.667
		(7.966)		(9.745)		(5.210)
Library exists	0.88	-0.017	0.967	-0.056	0.736	0.03
		(0.043)		(0.045)		(0.082)
Librarian exists	0.066	-0.03	0.087	-0.057	0.02	0.027
		(0.030)		(0.041)		(0.031)
Water	0.851	0.048	0.896	-0.004	0.771	0.139*
		(0.054)		(0.066)		(0.071)
Toilet	0.845	-0.021	0.902	-0.115	0.75	0.146**
		(0.055)		(0.074)		(0.071)
School reading score	-0.019	0.067	-0.001	0.061	-0.046	0.073
		(0.060)		(0.070)		(0.100)
Panel C: Baseline test						
Language	-0.046	-0.069	-0.144	0.031	0.109	-0.229**
		(0.081)		(0.106)		(0.100)
Math	-0.025	-0.085	-0.129	-0.007	0.141	-0.200**
		(0.074)		(0.094)		(0.098)
Total	-0.037	-0.088	-0.15	0.008	0.144	-0.233**
		(0.080)		(0.102)		(0.104)
Panel D: Individual characteristics						
Male	0.521	-0.008	0.536	-0.013	0.496	-0.004
		(0.017)		(0.021)		(0.027)
Hindu	0.797	0.027	0.867	-0.02	0.694	0.084
		(0.044)		(0.042)		(0.087)
Kannada mothertongue	0.629	0.008	0.695	-0.024	0.535	0.042
		(0.048)		(0.052)		(0.083)
Mothertongue/ medium different	0.239	-0.007	0.275	-0.032	0.185	0.024
		(0.032)		(0.042)		(0.049)
Scheduled Tribe/Caste	0.578	0.026	0.532	0.027	0.682	0.014
		(0.057)		(0.072)		(0.070)
Age	9.156	0.031	9.004	0.135*	9.385	-0.104
		(0.064)		(0.081)		(0.074)

Notes: Table compares probability of being tested in the follow-up (panel A) and baseline characteristics for all students who were tested at baseline but were not tested at follow-up (panels B-D). Comparisons are for the full sample (columns 1-2), hubs (columns 3-4) and spokes (columns 5-6). Columns 1, 3 and 5 show the mean of each variable for the control students. Columns 2, 4 and 6 show the difference in the mean of each variable for treatment and control students. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 5: Comparison of Non-Attriting Students

	All		Hubs		Spokes	
	C mean	T-C	C mean	T-C	C mean	T-C
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: School characteristics						
Number students	96.469	-6.851	109.579	-5.866	55.11	-5.105
		(7.386)		(8.593)		(6.228)
Library exists	0.95	-0.043	0.983	-0.032	0.847	-0.062
		(0.027)		(0.023)		(0.070)
Librarian exists	0.078	-0.036	0.076	-0.04	0.086	-0.026
		(0.032)		(0.035)		(0.067)
Water	0.886	0.012	0.903	-0.002	0.833	0.061
		(0.053)		(0.062)		(0.069)
Toilet	0.908	-0.06	0.936	-0.106*	0.818	0.079
		(0.045)		(0.055)		(0.064)
School reading score	0.043	0.032	0.06	0.011	-0.01	0.095
		(0.060)		(0.069)		(0.090)
Panel B: Baseline test						
Language	0.017	-0.003	-0.05	0.089	0.231	-0.284***
		(0.068)		(0.080)		(0.085)
Math	0.009	0.021	-0.064	0.113	0.239	-0.264***
		(0.072)		(0.083)		(0.094)
Total	0.014	0.015	-0.066	0.118	0.264	-0.303***
		(0.073)		(0.084)		(0.094)
Panel C: Individual characteristics						
Male	0.471	0.02	0.476	0.016	0.455	0.03
		(0.012)		(0.016)		(0.020)
Hindu	0.862	0.013	0.893	0.001	0.764	0.057
		(0.034)		(0.034)		(0.074)
Kannada mothertongue	0.683	-0.025	0.722	-0.06	0.565	0.081
		(0.044)		(0.047)		(0.075)
Mothertongue/ medium different	0.237	0.033	0.245	0.054	0.213	-0.021
		(0.033)		(0.041)		(0.051)
Scheduled Tribe/Caste	0.561	0.021	0.527	0.016	0.694	0.005
		(0.046)		(0.052)		(0.069)
Age	8.812	-0.02	8.853	-0.014	8.683	-0.026
		(0.042)		(0.047)		(0.080)

Notes: Table compares baseline characteristics for all students who were tested at baseline and were tested at follow-up. Comparisons are for the full sample (columns 1-2), hubs (columns 3-4) and spokes (columns 5-6). Columns 1, 3 and 5 show the mean of each variable for the control students. Columns 2, 4 and 6 show the difference in the mean of each variable for treatment and control students. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 6: Follow-up scores

	All		Hubs		Spokes	
	T-C	Controls	T-C	Controls	T-C	Controls
	(1)	(2)	(3)	(4)	(5)	(6)
Language	0.002	-0.004	0.063	0.028	-0.196	-0.115
	(0.074)	(0.070)	(0.080)	(0.077)	(0.131)	(0.116)
Math	-0.009	-0.016	0.008	-0.028	-0.076	-0.045
	(0.089)	(0.086)	(0.095)	(0.092)	(0.140)	(0.129)
EVS	-0.042	-0.05	-0.042	-0.075	-0.06	-0.019
	(0.082)	(0.079)	(0.090)	(0.089)	(0.129)	(0.111)
Total	-0.018	-0.026	0.017	-0.023	-0.139	-0.074
	(0.088)	(0.085)	(0.094)	(0.092)	(0.148)	(0.130)

Notes: Table shows difference in follow-up test scores between treatment and control students for the full sample (columns 1-2), hubs (columns 3-4) and spokes (columns 5-6). Columns 1, 3 and 5 show the raw differences in mean test scores between treatment and control students. Columns 2, 4 and 6 add controls for baseline differences that could affect follow-up test scores. Controls include normalized baseline test score, normalized baseline test score squared, school reading score (used in stratification) and indicators for grade, male, Hindu, Kannada, scheduled tribe/caste, age category and having a mother tongue different to the schools medium of instruction. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 7: Follow-up language scores by competency

	All		Hubs		Spokes	
	T-C	Controls	T-C	Controls	T-C	Controls
	(1)	(2)	(3)	(4)	(5)	(6)
Grammar	-0.063	-0.065	-0.018	-0.039	-0.221**	-0.167*
	(0.066)	(0.062)	(0.078)	(0.074)	(0.112)	(0.101)
Reading comprehension	-0.039	-0.044	0.009	-0.019	-0.196*	-0.124
	(0.063)	(0.061)	(0.068)	(0.068)	(0.116)	(0.106)
Punctuation	0.047	0.043	0.109	0.079	-0.109	-0.059
	(0.074)	(0.069)	(0.081)	(0.076)	(0.127)	(0.117)
Vocabulary	0.041	0.041	0.084	0.069	-0.114	-0.063
	(0.074)	(0.070)	(0.089)	(0.085)	(0.100)	(0.090)

Notes: Table shows difference in follow-up language test scores broken down by competencies between treatment and control students for the full sample (columns 1-2), hubs (columns 3-4) and spokes (columns 5-6). The competencies are based on broad categories from the tests. Columns 1, 3 and 5 show the raw differences in mean test scores between treatment and control students. Columns 2, 4 and 6 add controls for baseline differences that could affect follow-up test scores. Controls include normalized baseline test score, normalized baseline test score squared, school reading score (used in stratification) and indicators for grade, male, Hindu, Kannada, scheduled tribe/caste, age category and having a mother tongue different to the schools medium of instruction. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 8: Heterogeneity in follow-up language test scores

	All		Hubs		Spokes	
	C mean	T - C	C mean	T - C	C mean	T - C
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
	0	-0.004	-0.073	0.028	0.23	-0.115
		(0.070)		(0.077)		(0.116)
Panel B: Gender						
Males	-0.102	-0.009	-0.181	0.033	0.159	-0.132
		(0.074)		(0.080)		(0.136)
Females	0.091	0	0.025	0.024	0.289	-0.105
		(0.072)		(0.081)		(0.114)
Panel C: Grade						
Grade 3	0	0.014	-0.139	0.063	0.339	-0.129
		(0.092)		(0.107)		(0.140)
Grade 4	0	-0.001	-0.091	0.046	0.228	-0.141
		(0.084)		(0.095)		(0.137)
Grade 5	0	-0.025	0.002	-0.016	-0.01	-0.118
		(0.087)		(0.092)		(0.163)
Panel D: Baseline language test						
Baseline quartile 1	-0.462	0.058	-0.522	0.094	-0.172	-0.133
		(0.121)		(0.133)		(0.174)
Baseline quartile 2	-0.114	0.008	-0.155	0.027	0.023	-0.04
		(0.080)		(0.089)		(0.153)
Baseline quartile 3	0.152	-0.037	0.098	-0.015	0.308	-0.097
		(0.074)		(0.082)		(0.123)
Baseline quartile 4	0.408	-0.021	0.351	0.048	0.537	-0.164
		(0.068)		(0.080)		(0.120)
Panel E: Individual characteristics						
Hindu	0.026	-0.01	-0.054	0.026	0.315	-0.163
		(0.072)		(0.077)		(0.131)
Kannada	0.047	0.037	-0.02	0.064	0.305	-0.111
		(0.076)		(0.083)		(0.131)
Mothertongue/ medium different	-0.079	0.001	-0.201	0.057	0.34	-0.186
		(0.095)		(0.096)		(0.200)
Scheduled Tribe/Caste	-0.052	0.021	-0.166	0.02	0.288	-0.061
		(0.093)		(0.099)		(0.169)
Panel F: Baseline school characteristics						
Had no library before	0.268	-0.022	-0.239	0.38	0.45	-0.181
		(0.249)		(0.235)		(0.241)
Had no librarian before	-0.008	0.002	-0.102	0.057	0.299	-0.166
		(0.074)		(0.081)		(0.122)

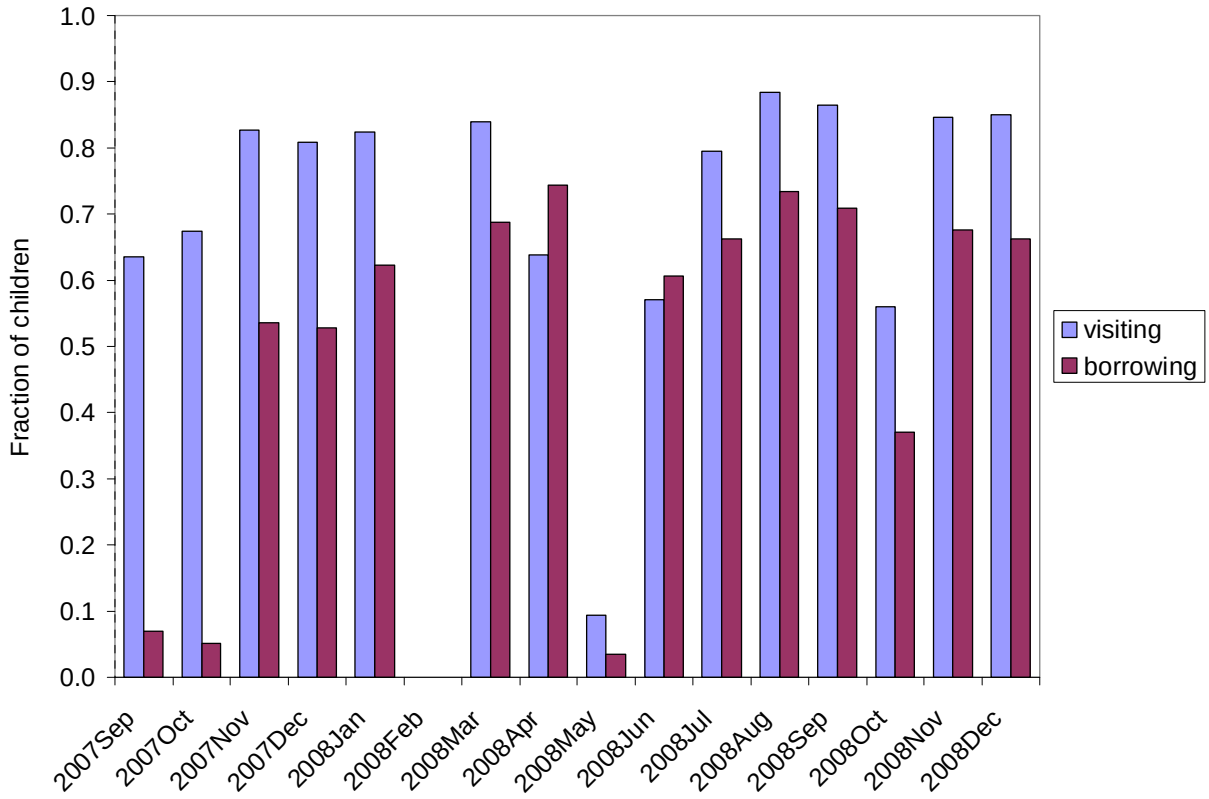
Notes: Table compares the aggregate follow-up scores for different subpopulations as indicated in each panel. Columns 1-2 are for the full sample, 3-4 are for the hubs, and 5-6 are for the spokes. Panel A represents the full sample for comparison For each of these subpopulations, columns 1, 3, and 5 list the mean of the control group and columns 2, 4, and 6 show the difference between the treatment mean and the control mean using the same controls as Table 6. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Table 9: Attendance rates

	All		Hubs		Spokes	
	C mean	T - C	C mean	T - C	C mean	T - C
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
	0.902	-0.002	0.9	-0.005	0.908	0.003
		(0.007)		(0.009)		(0.010)
Panel B: Gender						
Males	0.899	-0.005	0.897	-0.006	0.904	-0.003
		(0.007)		(0.009)		(0.010)
Females	0.906	0.001	0.903	-0.003	0.912	0.009
		(0.008)		(0.009)		(0.011)
Panel C: Grade						
Grade 3	0.901	0	0.898	-0.001	0.909	0.003
		(0.007)		(0.009)		(0.011)
Grade 4	0.9	-0.001	0.898	-0.005	0.904	0.009
		(0.009)		(0.010)		(0.011)
Grade 5	0.906	-0.006	0.904	-0.007	0.911	-0.003
		(0.009)		(0.011)		(0.011)
Panel D: Baseline language test						
Baseline quartile 1	0.885	-0.004	0.889	-0.013	0.873	0.021
		(0.010)		(0.012)		(0.015)
Baseline quartile 2	0.9	-0.001	0.897	-0.002	0.907	0.002
		(0.008)		(0.010)		(0.012)
Baseline quartile 3	0.911	-0.006	0.909	-0.008	0.917	-0.001
		(0.007)		(0.008)		(0.011)
Baseline quartile 4	0.912	0.003	0.907	0.003	0.922	0.008
		(0.008)		(0.010)		(0.010)
Panel E: Individual characteristics						
Hindu	0.909	-0.004	0.906	-0.006	0.915	0.003
		(0.007)		(0.009)		(0.009)
Kannada	0.916	-0.009	0.914	-0.013*	0.922	0
		(0.006)		(0.008)		(0.009)
Mothertongue/ medium different	0.88	0.011	0.875	0.014	0.897	0.002
		(0.014)		(0.017)		(0.017)
Scheduled Tribe/Caste	0.903	0	0.9	-0.004	0.913	0.007
		(0.008)		(0.010)		(0.012)
Panel F: Baseline school characteristics						
Had no library before	0.896	0.008	0.915	-0.02	0.89	0.019
		(0.022)		(0.032)		(0.026)
Had no librarian before	0.902	-0.001	0.9	-0.002	0.91	0.001
		(0.008)		(0.009)		(0.010)

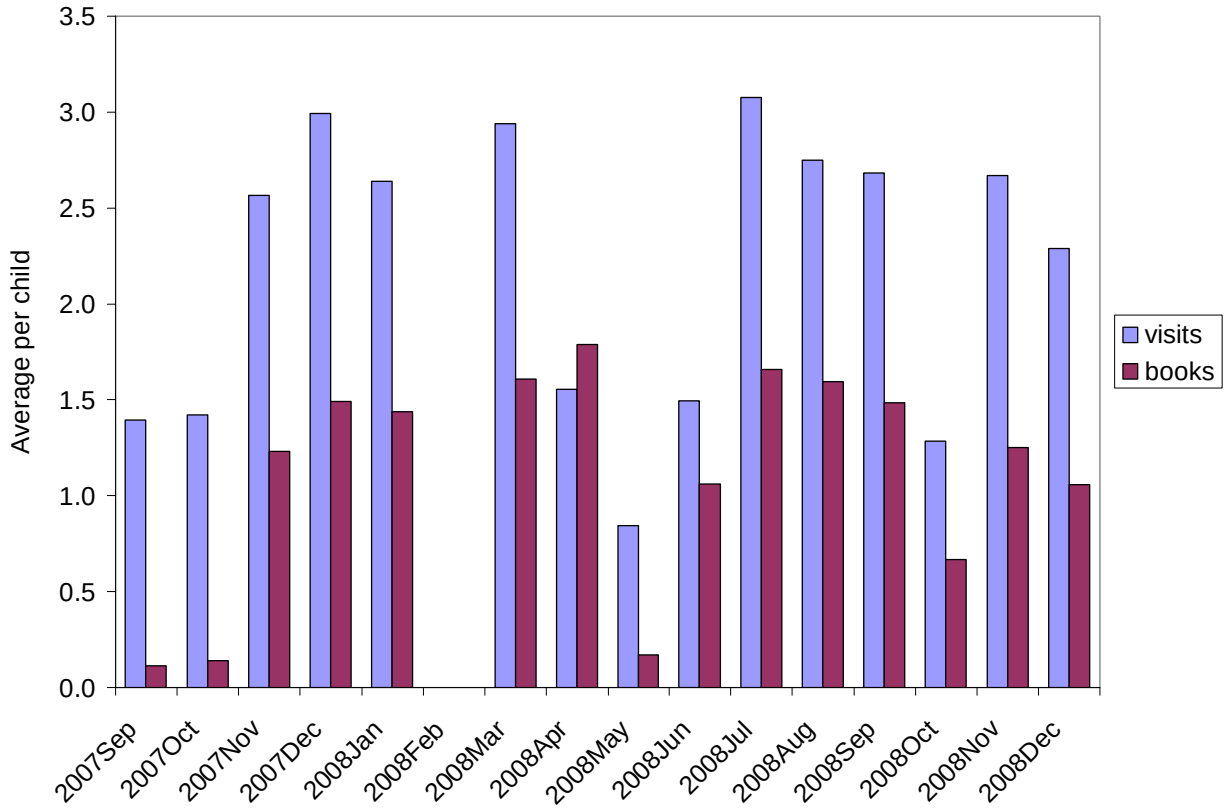
Notes: Table compares the attendance rates for different subpopulations as indicated in each panel. Attendance rate is defined as number of days a student is present over the number of days a school is open from September 2007 to March 2008. Columns 1-2 are for the full sample, 3-4 are for the hubs, and 5-6 are for the spokes. For each of these subpopulations, columns 1, 3, and 5 list the mean of the control group and columns 2, 4, and 6 show the difference between the treatment mean and the control mean. No controls are used in these differences. All observations are weighted by the number of classes in the relevant school-grade in order to ensure a representative sample.

Figure 1: Fraction of children visiting and borrowing in sample schools



Notes: Figure shows average fraction of children in grades 2-7 visiting the library at least once in the month during library period and average fraction borrowing at least one book in the month. Data are for all sample schools reporting in a particular month.

Figure 2: Average numbers of visits and books borrowed in sample schools



Notes: Figure shows unconditional average numbers of library period visits and books borrowed for children in grades 2-7. Data are for all sample schools reporting in a particular month.