

The Effects of School Libraries on Language Skills: Evidence from a Randomized Controlled Trial in India¹

March 2013

Evan Borkum
(Mathematica Policy Research)

Fang He
(US Government Accountability Office)

Leigh L. Linden
(The University of Texas at Austin)

Abstract: We conduct a randomized controlled trial of an Indian school library program. Overall, the program had no impact on students' scores on a language skills test administered after 16 months. The estimates are sufficiently precise to rule out effects larger than 0.053 and 0.037 standard deviations, based on the 95 and 90 percent confidence intervals. This finding is robust across individual competencies and subsets of the sample. The method of treatment, however, does seem to matter—physical libraries have no effect, while visiting librarians actually reduce test scores. We find no impact on test scores in other subjects or attendance rates.

JEL Codes: I21, I28, O15

Key Words: Library, Randomized Controlled Trial, Education, Development

¹ 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. Borkum: Mathematica Policy Research, P.O. Box 2393 Princeton, NJ 08543-2393, eborkum@mathematica-mpr.com. He: US Government Accountability Office, 441 G Street, NW, Room 4G48, Washington, DC 20548, hef@gao.gov. Corresponding Author. Linden: Department of Economics, The University of Texas at Austin, 1 University Station, BRB 1.116, C3100, Austin, Texas 78712, leigh.linden@austin.utexas.edu, phone: +1 (512) 495-8556, fax: +1 (512) 471-3510

Section I: Introduction

A quarter of all adults cannot read, and the vast majority of illiterate adults live in developing countries (UNESCO, 2011). As a result of striking deficits like these, development policy has increasingly focused on improving the quality of education. Recently, attention has focused on reading instruction, and in particular, on the availability of age-appropriate reading material in primary schools. In many countries, native language children's books simply do not exist, and even where they do, they are not available in schools.²

Unfortunately, our understanding of the education production process in the developing world is incomplete. Existing evidence on the effects of improving the availability of resources, like reading materials, in schools is mixed (Glewwe and Kremer, 2006). Some studies suggest that specific subgroups of students (such as high performing students) may benefit (Glewwe, Kremer, and Moulin, 2007), and there is also evidence that comprehensive programs that involve changes in instructors' teaching methods as well as additional resources work (for example, Banerjee et al., 2007). However, no consistent evidence exists that suggests students in general will benefit from greater access to books and other reading materials.

We study a promising educational program in Bangalore, India, that provides high quality libraries to public primary schools. The intervention provides children with access to well-equipped libraries, stocked with books supporting the school curriculum and staffed with a dedicated librarian. In addition to regulating access to the collection, the

² For example, USAID, WorldVision and AUSAid have recently announced the All Children Reading Grand Challenge with two main focuses – one of which is the availability of age- and language-appropriate reading material for primary school children. (All Children Reading, 2013)

librarian also provides regular reading-focused educational activities to encourage use of the library and facilitate interaction with the available books.

We evaluate the program using a randomized controlled trial (RCT) with a sample of 386 schools randomly selected from all public schools in Bangalore. Although all of the libraries functioned as expected and were utilized by a large number of the students (81 percent a month on average), we find no overall effect on students' language skills. The standard errors are small enough that we can rule out effects larger than 0.053 and 0.037 standard deviations based on the 95 and 90 percent confidence intervals. We also find no effect for most subsets of the sample, including divisions by grade, baseline test score, gender, and demographic characteristics. The one exception is by mode of implementation. We find no effect when the libraries are provided directly to schools, but we find a sizable negative effect if the libraries are provided through a visiting librarian. We also find no differences in students' scores on other subjects of the follow-up exam or in students' average attendance rates. These results suggest that providing access to reading material may not be sufficient to improve students' language skills.

This study also directly builds on an existing literature that investigates the education production function for reading in developing countries. For example, Banerjee et al. (2007) find that low-cost remedial programs using individuals from the local communities to teach basic literacy skills to young children substantially improved performance. He, Linden, and MacLeod (2009) show that improved teaching practices have a significant effect on young students reading skills. And Abeberese et al. (2010) find significant effects on students' reading scores of a Philippine read-a-thon program.³

³ There is also related work on reading programs in the developed world including Machin and McNally (2008), Fryer (2011), and Kim and Guryan (2010). Researchers in other fields have also contributed to this

The remainder of the paper is set out as follows. Section II describes the library program in detail. Section III discusses the research design. Section IV verifies the internal validity of the study, and Section V estimates the effects of the library program. 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 organized according to a hub and spoke system with each hub school attached to several spokes in the same geographic 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 are recruited externally by Akshara and undergo several group-training sessions after recruitment as well as several refresher sessions during the academic year.

Training focuses on library operations and on educational activities. The librarian’s role includes maintaining the library, issuing books and keeping borrowing records. Akshara provides a room full of age and language appropriate material, designed to support the existing curriculum. 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 may select their own books within a given level to borrow). The activities include storytelling, role-playing games

literature (e.g. Kim, 2007; and 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).

(where children act out a story from a book) and other educational games (such as identifying the sounds made by animals in a story book). Librarians conduct these activities for each class during regularly scheduled library periods. These 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 on a regular schedule.⁴ The mobile librarian transports books from the hub library and issues them to the children using the same color-coded system as in the hub libraries. Typically, the mobile librarian will spend several hours at a spoke school, serving students by class. Unlike the hub libraries, the librarians did not conduct activities in the spoke schools for most of the study period, although they did introduce these activities in the final three months of the study.

Hub and spoke facilities differ dramatically from the resources they replace.⁵ Most primary schools do have a small collection of books. However, the books are rarely age and language appropriate. The collections are very limited. Many (57 percent) are kept in a cupboard – almost all of which (92 percent) are located in the headmaster’s office – and although children and teachers are supposed to have access to the material, they rarely do. Only 20 schools (5 percent of the sample) designate someone to facilitate use of the books, and even then, they are not dedicated exclusively to the task. Only 4 schools (1 percent of the sample) have both a dedicated room and someone designated to help children and teachers work with the available books.

⁴ 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.

⁵ The averages presented in this paragraph are for the entire sample of schools from the school-level baseline. Estimates for the treatment and control groups are provided in Table 2.

Data collected during the experiment allows us to document how the libraries were used by students and to confirm that students interacted with the libraries at the intended levels of intensity. The average fraction of children visiting the library in each month of the study fluctuates between 50 and 90 percent. The exceptions are February 2008 (school exams) and May 2008 (summer vacation), but overall 81 percent of the children visit the libraries at least once a month during the academic year. Borrowing rates are similarly high. In most months, over 60 percent of students borrow at least one book. And excluding months during exams and holidays, children in the average school visit the library 2.4 times and borrow 1.3 books a month. Usage is 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.⁶

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. First, Akshara arranged all government primary schools in Bangalore into a hub and spoke network. They chose hubs based on size, geographic location and the availability of a room to house a library. The remaining schools became spokes of a nearby hub, with each hub attached to up to seven spoke schools. Given significant variation in school size, we excluded from the research sample

⁶ 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.

the upper and lower 10 percent of spoke schools (removing those with enrollment of less than 20 or greater than 230) and the smallest 5 hub schools, which had total enrollment of 17 students or less. We then randomly selected a single spoke for each hub (we refer to the pair as a “unit”), yielding a sample of 300 units.⁷

Through power calculations, we determined that only 200 units would be included in the study.⁸ 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 we first grouped units by geographic location (we used “blocks” which are most analogous to US zip codes), ranked them by the average test score of the unit, and then grouped the units into triplets.⁹ Within each triplet, we selected one unit for the treatment group, one for the control group, and one to be left out of the study, all with equal probability.^{10,11} For each school in the sample, we then randomly chose a single class from grades three, four, and five.

B. Data Collection

We use three original sources of data. First, we have two data sets that provide baseline information: a school-level census conducted at the end of the 2006-7 academic year and a student-level survey of the sample schools collected at the beginning of the 2007-8 academic year. Since the randomization occurred after the 2006-7 academic year, the

⁷ Although only one spoke was included in the analysis, all spokes still received the intervention. Additionally, the elimination of the smallest 10 percent of spoke schools from the research sample resulted in a small number of hubs (those with only to small spokes) not having an assigned spoke in the data set.

⁸ We chose the sample to be large enough to detect a minimum effect size of 0.15 standard deviations with 90 percent power at the 5 percent significance level.

⁹ We grouped unmatched units into triplets by average test score irrespective of geographic location.

¹⁰ We did not employ a strategy that utilized re-randomization if the resulting research groups were unbalanced.

¹¹ In results available upon request, we have verified that the selected sample is comparable to the schools that were omitted from the study.

school-level baseline provides data from before the randomization while the student level-baseline provides data from after randomization but before the libraries began operating.¹² We then conducted a student-level follow-up survey in December of the 2008-9 academic year, just before schools administered end-of-year exams.

The school-level baseline data set includes every school in the city. It provides information on the physical characteristics of the school, as well as average reading test scores for all students. These were used to stratify the sample for the randomization and to construct the hub and spoke networks.

The student-level baseline provides a quick gauge of students' general aptitude. 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 much more comprehensive. 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 for the state's three primary subjects: math, language arts, and science (referred to as environmental science (EVS)). We then piloted the questions and created a separate exam for each grade by eliminating the questions that yielded little variation in student performance.¹³ The math and science tests allow us to

¹² The disadvantage of this strategy is that the composition of students may have changed after the randomization. However, we can directly test for this possibility. Specifically, we demonstrate that the research groups were balanced based on information collected prior to the randomization using the school-level data (Table 2), and then confirm that the groups were balanced at the start of the next school year using the student-level baseline (Table 3). The results are consistent with other studies with similar survey schedules (see for example, Banerjee et al., 2009). Student enrollment does not seem to respond to the potential receipt of such programs.

¹³ Because many children perform below grade level, we could not use all of the competencies in each 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.

assess whether or not better language skills translate into better performance in other subjects or whether the use of the libraries came at their expense.¹⁴

Finally, we collected two types of administrative information from the schools. The first consisted of basic demographic information from the school registers. This included students' gender, age, language at home, religion and caste. We also collected daily attendance for all children from the class rosters. Such data must always be used with caution since teachers may incorrectly record attendance data when they have an incentive to do so (Linden and Shastry, 2012). However, in this context, the intervention does not change teachers' incentives to report students as being either absent or present.

While the school census data is available for all schools in the sample, 16 schools refused to participate in the student-level data collection processes despite prior agreements with both the schools and administrators. However, given the small number of schools, this did not create an imbalance in the characteristics of the treatment and control groups.¹⁵ For the student-level data, this yields a sample of 370 schools (193 hubs and 177 spokes) and 20,858 students (14,455 in hub schools and 6,403 in spoke schools). Table 1 contains a full tabulation of the sample.

¹⁴ 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.

¹⁵ Compared to the overall sample of 386 schools, this represents a loss of only 4.2 percent of the sample. The danger, like the danger with all attrition, is that differences in the attrition patterns between treatment and control schools could make the two research groups incomparable. Given the small number of schools, this is very unlikely. However, estimates using the census data, which are available on all school in the sample, we have verified that the loss of these schools did not unbalance the sample (Table 2). This is also born out in Table 3 which demonstrates that student characteristics are balanced for the 95.8 percent of schools remaining in the sample.

C. Analytic models

The basic research design compares students in the treatment and control groups, and the randomization ensures that the assignment of schools to the treatment is independent of school and student characteristics. As a result, we can directly estimate the effects of the intervention by comparing the average outcomes in the treatment and control groups using the following equation:

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

where Y_{ijkl} is the outcome for student i in grade j , school k , and unit l . The variable $Treat_l$ is an indicator for whether unit l is assigned to the treatment group. We also include a vector of control variables X_{ijkl} . This includes baseline test scores by subject, average reading scores of the hub-spoke unit from the school-level baseline, gender, grade, age, majority religion and language group affiliations, caste status and an indicator whether the language a student speaks at home differs from the medium of instruction.¹⁶ Our preferred specification also includes fixed effects for the triplets used to stratify the schools for randomization. Finally, we cluster standard errors at the level of treatment assignment – the hub and spoke unit.

Section IV: Internal Validity

If the treatment assignment is statistically independent of the characteristics of the schools and the students, then the treatment and control students should, on average, have

¹⁶ 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.

similar characteristics. We check this using the demographic data and the school- and student-level baselines in Tables 2 and 3.

Starting with Table 2, Panel A includes all sample schools that were included in the randomization. The mean values from the schools assigned to the control group are in the first column while the second column contains the estimates differences using equation (1) without any control variables. Of the 10 variables tested,¹⁷ only one is statistically significant at conventional levels (Collection Kept in Separate Room) and it is practically small. Overall, the joint test cannot reject equality of the averages across all variables at conventional levels of significance (the p-value is 0.416). This confirms that the randomization created comparable treatment and control groups.

After the randomization, however, 16 schools refused to participate in the student-level data collection. While this only represents 4.1 percent of the original sample, we confirm in Panel B that their loss did not make the treatment and control groups dissimilar. All of the estimates resemble those in Panel A, and the joint test is not significant at conventional levels (the p-value is 0.398). As a result, we conclude that the loss of the 16 schools did not compromise the internal validity of the study.

Next, we turn to the student-level estimates in Table 3. Panel A contains estimated differences using all students surveyed at the start of the 2007-8 academic year, while Panel B includes just those students who also completed the follow-up test. Panel A confirms the comparability of the research groups at the start of the study. Of the 13 differences tested, only one is statistically significant at conventional levels (Grade), and

¹⁷ Note that in the tested variables we omit corresponding categories if they are the exact complement to a provided measure. In Table 2, for example, we omit Urdu medium schools because the schools in our sample are either Kannada or Urdu. The differences for the two categories are perfectly correlated.

the 3.1 percentage point difference is practically small. The overall joint test is also not statistically significant (the p-value is 0.517).

Finally, we check in Panel B to ensure that student-level attrition did not compromise the internal validity of the study. As shown in row one, 72.3 percent of the students who completed the baseline test in the control group also completed the follow-up exam, and the average rate of completion in treatment group was only 0.7 percentage points less. The other estimates in the panel are consistent with those in Panel A, except that the difference in the average grade level is no longer statistically significant.

Section V: Treatment Effects

A. Language Scores

Having verified the internal validity of the experiment, we begin by estimating the effects on students' language test scores in Table 4. Panel A contains the estimated treatment effects on the overall score. In columns one, two and three, we estimate the effects using equation (1) without and with control variables. The estimates are consistent with each other, which further emphasizes the balance in characteristics between the research groups. However, the estimates are small and statistically insignificant at conventional levels. These suggest that the provision of the libraries had no effect on students' test scores. With the full set of controls, we can reject effects larger than 0.053 and 0.037 standard deviations. Finally, in column four, we add classroom random effects in an attempt to further improve precision, but the confidence intervals remain unchanged.

Recently, many viable interventions have yielded effects in the range of 0.12 to 0.70 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 comfortably reject the hypothesis that the treatment effect of the libraries is in the same range of these successful interventions.

While the program may not have had an effect on test scores overall, it may have still affected individual competencies. In Panel B, we estimate treatment effects for the four competencies that were included on most versions of the test. The results are remarkably similar to those in Panel A. For no specification do we find a positive, statistically significant effect on students' test scores for any subject. As a result, we conclude that the libraries had no overall effect on students' language skills.

B. Other Subjects

We test the effects of the program on other subjects in Table 5. The table is organized like Table 4, but with the math score effects in Panel A and the science effects in Panel B. For both subjects we find no consistent evidence of positive or negative spillovers to other subjects. In our preferred specification (column 3), we actually find a statistically significant negative effect on science scores at the ten percent level, but none of the other estimates are statistically significant, and the estimate for math using the preferred specification is also not statistically significant. We conclude that like the effects on language skills, the program had no effect on other subjects.

C. Language Scores: Heterogeneous Effects

Despite the lack of an effect for the full sample, subgroups may have still benefitted. Using our preferred specification, we re-estimate the impact of the program on the follow-up language test score in Table 6 by dividing our sample both by mode of implementation – hub or spoke – and by individual student characteristics.¹⁸ For each group of schools in the indicated columns, Panel A provides the estimated treatment effect for all students and Panels B-E contain the estimates divided by gender, grade, baseline test score, and other student-level characteristics.

The only variation we observe is by mode of treatment.¹⁹ The hub schools have no effect on students test scores. And like the estimates for all schools, this is consistent across all subgroups of students. The only exceptions are scheduled caste or tribe students in the overall sample, students attending a school whose medium is different than the language they speak at home for the hub schools, and students in the top quartile of the baseline test distribution for the hubs. However, these are only three of 28 estimates.

We do, however, find a large, significant negative effect for the spoke schools. Like the other group, the effects are remarkably consistent across the subgroups. Why the spoke treatment would negatively affect students is unclear. The primary difference between the modalities is that, for the spoke schools, the librarian visited the schools on a pre-arranged schedule and could only interact with students during that time. One possibility is that these visits disrupted the normal school schedule, and teachers adjusted

¹⁸ In results available upon request, we also estimate the effects using school characteristics from the school census, but find the same pattern of effects as observed in Table 6.

¹⁹ In results not included in this article, we have also disaggregated the effects on individual competencies listed in Panel B of Table 4 and the scores on other subjects by mode of implementation. For the individual competencies, the results are similar to those presented in Table 6 with the hub treatment having no effect and the spoke treatment creating sizable negative effects. For the other subjects, however, we find no statistically significant effect for either treatment modality.

by reducing the time spent on language arts.²⁰ Hub schools teachers, on the other hand, had much more scheduling flexibility and could have chosen the least disruptive times.

D. Attendance

Even without a positive effect on test scores, the libraries could affect student attendance – for example, the library could provide an incentive if students enjoyed the visits. To assess these effects, we estimate equation (1) with the attendance rate as the dependent variable and the full set of controls. The attendance rate is the fraction of all days on which the school was open that a student was marked present. The results are presented in Table 7. The overall attendance rates are fairly high: the means for the control 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 no impact on attendance rates. This is consistent with other studies that find no effect on attendance (Banerjee et al., 2007; He, Linden and MacLeod, 2008).

Section VI: Conclusion

This study estimates the effects of providing students with reading material through school libraries. The program provides high-quality reading material designed to support the existing language curriculum as well as the support of a dedicated librarian. Overall the treatment had no effect on students' language skills or on their performance in other subjects. The results are remarkably consistent across individual language competencies

²⁰ The estimates for the effects on other subjects by mode of treatment are not statistically significant for either subject in both modalities. Estimates are available upon request.

and also across individual subsets of the student population. We can reject treatment estimates larger than 0.053 and 0.037 at the 95 and 90 percent confidence levels, using our preferred specification. We also find that the method of providing the library resources matters. Placing a library in a school has no effect, while providing resources through a mobile librarian actually causes students to learn less.

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 show large positive effects of programs that provide additional resources while also changing the way that students are taught during the normal school day. Our results are consistent with studies that find the provision of resources alone insufficient to improve student performance (Glewwe and Kremer, 2006). This is not to suggest that such programs are valueless. It is possible that programs such as this might enhance the effects of other interventions. For example, changes in the pedagogical strategies for teaching reading may be more effective if students have access to significant reading material. But the results do indicate that understanding these linkages is an area for future study.

References

- Abeberese, A., T. Kumler, and L. Linden (2011) "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.
- All Children Reading (2013) "Teaching and Learning Materials," <http://www.allchildrenreading.org/teaching-and-learning-materials>. Accessed February 10, 2013.
- 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.
- 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.
- 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, 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.
- 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.
- Linden, L. and G. Shastry (2012) "Identifying Agent Discretion: Exaggerating Student Attendance in Response to a Conditional School Nutrition Program," *Journal of Development Economics*. 99(1): 128-138.
- 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 Tabulation

	Control	Treatment	All
	(1)	(2)	(3)
School-Level Baseline			
Schools	195	191	386
Student-Level Baseline			
Schools	188	182	370
Classes	553	541	1,094
Students	10,960	9,898	20,858

Notes: Table shows composition of sample for both the school-level survey conducted before the randomization and the student-level baseline conducted after the randomization but prior to the start of the intervention.

Table 2: Internal Validity, School-Level Baseline

	Control			Control	
	Mean	Difference		Mean	Difference
	(1)	(2)		(3)	(4)
Panel A: Full Sample			Panel B: Participating in Student-Level Baseline		
Reading Test Score	0.053	0.002 (0.054)	Reading Test Score	0.052	0.009 (0.056)
Number students	214.519	-28.045 (22.148)	Number students	217.796	-32.041 (22.566)
Water	0.862	0.031 (0.039)	Water	0.868	0.02 (0.040)
Toilet	0.855	-0.023 (0.040)	Toilet	0.866	-0.042 (0.041)
Kannada Medium	0.913	-0.028 (0.030)	Kannada Medium	0.91	-0.019 (0.031)
Book Collection	0.865	-0.033 (0.037)	Book Collection	0.871	-0.036 (0.038)
Collection Kept in Separate Room	0.147	-0.066** (0.032)	Collection Kept in Separate Room	0.142	-0.063* (0.032)
Collection Kept in Cupboard	0.574	-0.009 (0.055)	Collection Kept in Cupboard	0.569	0.002 (0.057)
Collection Kept in Head Master's Office	0.528	-0.002 (0.055)	Collection Kept in Head Master's Office	0.527	0.009 (0.057)
Designated Staff Member	0.059	-0.011 (0.024)	Designated Staff Member	0.061	-0.011 (0.025)
Joint Test			Joint Test		
Chi ² (10)		10.29	Chi ² (10)		10.49
P-Value		0.416	P-Value		0.398

Notes: Table compares schools assigned to the treatment and control groups using the school-level baseline conducted prior to the randomization. The sample in Panel A includes all 200 schools included in the randomization and assigned to either the treatment or control group. Panel B includes just those schools that participated in the student-level baseline and follows-up surveys, omitting the 16 schools that refused to participate. Columns one and three provide the mean for the control group while columns two and four provide the estimated difference between the two groups using equation (1) without any control variables. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.

Table 3: Internal Validity, Student-Level Baseline

	Control			Control	
	Mean	Difference		Mean	Difference
	(1)	(2)		(3)	(4)
Panel A: Completing Baseline Survey			Panel B: Completing Baseline and Follow-Up		
Language Arts	0.018	-0.048 (0.066)	Took Follow-Up Test	0.723	-0.007 (0.018)
Math	0.022	-0.055 (0.066)	Language Arts	0.036	-0.029 (0.068)
Total	0.022	-0.058 (0.069)	Math	0.035	-0.03 (0.069)
Grade	4.001	0.031** (0.015)	Total	0.039	-0.033 (0.072)
Male	0.483	0.006 (0.010)	Grade	3.956	0.015 (0.024)
Hindu	0.778	0.02 (0.034)	Male	0.473	0.011 (0.011)
Muslim	0.125	-0.004 (0.029)	Hindu	0.81	0.027 (0.033)
Kannada at Home	0.568	-0.015 (0.037)	Muslim	0.111	0.001 (0.028)
Urdu at Home	0.127	-0.006 (0.029)	Kannada at Home	0.589	-0.021 (0.037)
Other Language at Home	0.146	0.025 (0.024)	Urdu at Home	0.114	-0.005 (0.028)
Home Language Different From Instruction Medium	0.2	0.013 (0.024)	Other Language at Home	0.149	0.035 (0.026)
Scheduled Tribe/Caste	0.289	0.003 (0.031)	Home Language Different From Instruction Medium	0.203	0.022 (0.026)
Age	8.889	-0.007 (0.040)	Scheduled Tribe/Caste	0.304	0.008 (0.032)
			Age	8.791	-0.015 (0.043)
Joint Test			Joint Test		
Chi2(13)		12.13	Chi2(14)		11.94
P-Value		0.517	P-Value		0.611

Notes: This table compares students in schools assigned to the treatment and control groups using the student-level baseline conducted at the start of the 2007-8 academic year, after the randomization and prior to the start of the treatment. Panel A includes all students completing the baseline test while Panel B includes just those students who completed both the baseline and follow-up exams. Columns one and three provide the mean for the control group while columns two and four provide the estimated difference between the two groups using equation (1) without any control variables. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.

Table 4: Effect on Language Test Scores

	(1)	(2)	(3)	(4)
Panel A: Language Skills				
Normalized Score	-0.024 (0.074)	-0.045 (0.063)	-0.047 (0.051)	-0.044 (0.049)
95 Percent Confidence Interval	(-0.171, 0.123)	(-0.169, 0.079)	(-0.147, 0.053)	(-0.141, 0.053)
90 Percent Confidence Interval	(-0.147, 0.099)	(-0.149, 0.059)	(-0.130, 0.037)	(-0.125, 0.037)
Panel B: Disaggregated by Competency				
Grammar	-0.087 (0.064)	-0.096 (0.060)	-0.090* (0.051)	-0.071 (0.050)
Reading comprehension	-0.052 (0.065)	-0.065 (0.056)	-0.051 (0.046)	-0.043 (0.045)
Punctuation	0.029 (0.072)	0.015 (0.062)	-0.014 (0.049)	-0.025 (0.047)
Vocabulary	0.009 (0.072)	-0.012 (0.061)	-0.03 (0.047)	-0.054 (0.044)
Model	OLS	OLS	OLS	Classroom Random Effects
Clustered by Unit	Yes	Yes	Yes	Yes
Control Variables	No	Yes	Yes	Yes
Triple Fixed Effects	No	No	Yes	Yes

Notes: This table presents estimates of the effects of the library intervention on students' language test scores. Panel A contains estimates of the effect on overall scores, while Panel B presents estimates of the effects on individual competencies. Estimates in column one contain no control variables. Those in column two include controls, and in column three, we add fixed effects for the individual triplet groups used to stratify the randomization. Column four includes all of the control variables and classroom level random effects. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.

Table 5: Follow-up scores

	(1)	(2)	(3)	(4)
Panel A: Math				
Normalized Score	-0.031 (0.087)	-0.065 (0.073)	-0.052 (0.061)	-0.034 (0.056)
95 Percent Confidence Interval	(-0.202, 0.140)	(-0.210, 0.080)	(-0.173, 0.069)	(-0.143, 0.076)
90 Percent Confidence Interval	(-0.174, 0.112)	(-0.187, 0.056)	(-0.153, 0.049)	(-0.126, 0.058)
Panel B: Science (EVS)				
Normalized Score	-0.047 (0.079)	-0.082 (0.067)	-0.100* (0.054)	-0.046 (0.051)
95 Percent Confidence Interval	(-0.202, 0.108)	(-0.214, 0.050)	(-0.207, 0.006)	(-0.146, 0.055)
90 Percent Confidence Interval	(-0.177, 0.083)	(-0.193, 0.028)	(-0.189, -0.011)	(-0.130, 0.039)
Model	OLS	OLS	OLS	Classroom Random Effects
Clustered by Unit	Yes	Yes	Yes	Yes
Control Variables	No	Yes	Yes	Yes
Triple Fixed Effects	No	No	Yes	Yes

Notes: This table presents estimates of the effects of the library intervention on students' math and science test scores. Panel A contains estimates of the effect on math scores, while Panel B presents estimates of the effects on science scores. Estimates in column one contain no control variables. Those in column two include controls, and in column three, we add fixed effects for the individual triplet groups used to stratify the randomization. Column four includes all of the control variables and classroom level random effects. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.

Table 6: Disaggregated Effects on Language Test Scores

	All		Hubs		Spokes	
	Control		Control		Control	
	Mean	Difference	Mean	Difference	Mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Students						
All Students	0.017	-0.047 (0.051)	-0.029	0.045 (0.058)	0.138	-0.218** (0.088)
Panel B: Gender						
Males	-0.082	-0.038 (0.056)	-0.129	0.06 (0.063)	0.046	-0.227** (0.101)
Females	0.109	-0.061 (0.050)	0.066	0.026 (0.058)	0.219	-0.232*** (0.087)
Panel C: Grade						
Grade 3	0.047	0.013 (0.072)	-0.047	0.143 (0.087)	0.244	-0.176* (0.103)
Grade 4	-0.002	-0.067 (0.061)	-0.062	-0.008 (0.079)	0.134	-0.262*** (0.096)
Grade 5	0.005	-0.064 (0.068)	0.02	-0.008 (0.069)	-0.061	-0.09 (0.264)
Panel D: Baseline language test						
Baseline quartile 1	-0.392	-0.024 (0.080)	-0.432	0.02 (0.097)	-0.274	-0.266* (0.153)
Baseline quartile 2	-0.09	-0.07 (0.059)	-0.123	0.008 (0.066)	0.006	-0.264** (0.124)
Baseline quartile 3	0.138	-0.044 (0.056)	0.094	0.012 (0.069)	0.25	-0.222** (0.103)
Baseline quartile 4	0.402	0.019 (0.059)	0.368	0.156* (0.081)	0.477	-0.148 (0.108)
Panel E: Individual characteristics						
Hindu	0.035	-0.084 (0.053)	-0.013	0.012 (0.057)	0.18	-0.268*** (0.099)
Kannada at Home	0.078	-0.017 (0.052)	0.039	0.058 (0.058)	0.194	-0.161 (0.109)
Home Language Different From Instruction Medium	-0.053	-0.041 (0.099)	-0.118	0.166* (0.100)	0.161	-0.435** (0.202)
Scheduled Tribe/Caste	-0.031	-0.170*** (0.063)	-0.14	-0.064 (0.073)	0.229	-0.318** (0.158)

Notes: This table presents estimates of the effects of the library intervention on students' language test scores disaggregated by mode of implementation and by socio-demographic controls. All differences are estimated using equation (1) with all controls and triplet fixed effects. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.

Table 7: Effect on Attendance

	All		Hubs		Spokes	
	Control		Control		Control	
	Mean	Difference	Mean	Difference	Mean	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All Students						
All Students	0.917	0 (0.004)	0.915	-0.002 (0.005)	0.924	0.003 (0.006)
Panel B: Gender						
Males	0.914	-0.002 (0.004)	0.911	-0.003 (0.005)	0.921	0.001 (0.007)
Females	0.921	0.002 (0.004)	0.918	-0.001 (0.005)	0.927	0.005 (0.007)
Panel C: Grade						
Grade 3	0.918	0 (0.004)	0.914	0.001 (0.006)	0.927	0.006 (0.007)
Grade 4	0.917	0.004 (0.005)	0.913	0.001 (0.006)	0.925	0.003 (0.007)
Grade 5	0.917	-0.002 (0.005)	0.917	-0.004 (0.006)	0.919	0.035*** (0.008)
Panel D: Baseline language test						
Baseline quartile 1	0.907	0.002 (0.006)	0.905	-0.004 (0.007)	0.914	0.022** (0.009)
Baseline quartile 2	0.915	0.002 (0.004)	0.913	0.002 (0.005)	0.922	0 (0.008)
Baseline quartile 3	0.92	-0.001 (0.004)	0.919	-0.003 (0.006)	0.925	-0.002 (0.007)
Baseline quartile 4	0.927	0.003 (0.005)	0.923	0.001 (0.006)	0.934	0.01 (0.009)
Panel E: Individual characteristics						
Hindu	0.92	-0.002 (0.004)	0.918	-0.004 (0.005)	0.928	0.006 (0.007)
Kannada at Home	0.924	-0.002 (0.004)	0.922	-0.005 (0.005)	0.931	0.005 (0.007)
Home Language Different From Instruction Medium	0.903	-0.009 (0.006)	0.899	-0.005 (0.008)	0.916	0 (0.028)
Scheduled Tribe/Caste	0.919	-0.002 (0.004)	0.915	-0.004 (0.006)	0.927	0.008 (0.008)

Notes: This table presents estimates of the effects of the library intervention on students' attendance rates disaggregated by mode of implementation and by socio-demographic controls. All differences are estimated using equation (1) with all controls and triplet fixed effects. Standard errors are clustered at the unit level. Significance at the one, five, and ten percent levels are indicated by *, **, and ***, respectively.