A Better Way to Teach Children to Read? Evidence from a Randomized Controlled Trial[1]

May 2009

Fang He
(Columbia University)

Leigh L. Linden[2]
(Columbia University, BREAD, and IZA)

Margaret MacLeod
(Columbia University and U.S. Department of State)

Abstract: Using three randomized evaluations, we assess the relative efficacy of a literacy skills development program in Mumbai, India on children in three types of educational institutions: public schools, a stand-along reading class, and pre-schools. The new methods prove effective in all populations and, on average, all implementation strategies yield gains of between 0.12 and 0.70 standard deviations in performance on a basic literacy assessment. The program seems to be more effective as a supplement to existing instruction rather than as a primary means of instruction. We find, for example, that when implemented in public schools, an out-of-school time version of the program significantly improves scores over an in-school version of the program. Additionally, we find that programs based in existing institutions (public and pre- schools) provide much more robust gains in student performance than the stand alone classes. Finally, in one experiment, we are also able to disaggregate the effect of the techniques by the reading ability of children's parents and find that while parents' skills are correlated with children's, the effects of the program are fairly uniform across all students.

JEL: C93, I21, I28
Key Words: Education, Randomized Controlled Trial, Early Literacy

---

[1] Projects of this magnitude require the contributions of a large number of people. While it would be impossible to recognize everyone who made a contribution to this project, we are particularly indebted to the efforts of Pratima Bandekar, Rajashree Kabare, Vidya Tiware, and Nandit Bhatt. Funding for this project was provided by the Pratham Mumbai Educational Trust and ICICI Bank of India. Finally, this project simply would not have happened without the guidance and support of Madhav Chavan and Rukmini Banerji . The views expressed in this paper are those of the authors, and do not necessarily reflect those of the U.S. Department of State or the U.S. Government.
[2] Corresponding author. Please direct all correspondence to leigh.linden@columbia.edu.

I. Introduction

The Millennium Development Goals and the Education for All initiatives have placed strong emphasis on the promotion of universal primary education. However, enrolling children is only a first step in educating children in developing countries. The quality of education also matters. As Hanushek and Woessmann (2007) document in their survey of the evidence, education quality differs significantly across countries and is especially low in developing countries. Consequently, to close the gap between developed and developing countries, programs to improve education quality are crucial. Effective methods to teach reading skills are particularly important. Existing research strongly suggests that learning to read at an early age lays a strong foundation for future academic skill development (Scarborough, 2001), suggesting that improvements in early reading programs could yield significant dividends.

This study tries to help answer this question by looking at a particular education program that seeks to improve the quality of language instruction for early learners in the slum communities of Mumbai, India. In this particular setting, the quality of language instruction is generally poor and the need to improve existing education models is critical, especially at an early age. The main purpose of the curriculum is to provide basic literacy training to pre- school children to lay a foundation for their matriculation to regular primary schools. However, as we show, the program also provides a useful intervention for early primary students because unlike the current teaching practices that focus on rote memorization and repetition, the *Shishuvachan* curriculum focuses on comprehension facilitated by teacher-student interaction centered around storytelling and classroom games.

The purpose of this study is to assess the effectiveness of the *Shishuvachan* curriculum. Recent studies have suggested that the efficacy of educational strategies may vary by educational institution and by type of child (Duflo, Dupas, and Kremer, 2008; He, Linden, and Macleod,

2008). As a result, we conduct three separate experiments over three years designed to evaluate the effects of the program on three different student populations in three different types of institutions: first-grade students in government schools, pre-school students in specially designed community-based *Shishuvachan* classes, and existing pre-school classes. In addition, for the government school population, we test two variants of the program – one that integrates the program into the regular school day and one that operates outside of school hours as a supplement.

Overall, we find that the program is quite effective an improving students reading scores. The effectiveness of the program on students within the individual institutions varies from 0.26 standard deviations to 0.7 standard deviations. The program seems to be more effective with students in certain institutions, proving most effective in the pre-schools classes for which it was originally designed. In addition, the program is particularly effective for low performing students. Additionally, the program seems to generate much more robust gains in student achievement when run as a complement to existing institutions rather than when run in a standalone, dedicated community based class. Finally, within schools, the out-of-school time model proves significantly more beneficial than the in-school model increasing test scores by 0.24 standard deviations beyond the 0.26 standard deviation effect of the within school model. This suggests that the program is most effective as a complement to teachers' existing techniques rather than as a pure substitute.

In our experiment with the existing pre-schools we also collect information on parents' reading abilities. As expected, students' reading skills are tightly correlated with parents' skills. Children of parents who have achieved the ability to read a paragraph completely score 0.36 standard deviations higher on our reading assessment than those whose parents cannot read a

paragraph. However, despite these initial differences in the study population, the teaching methods seem to have similar effects on all students.

This study complements a recent push in economics for randomized evaluations of education programs in developing countries (for a review see Kremer (2003) and for a discussion on randomized evaluations in developing countries see Duflo, Glennerster, and Kremer (2007)). Among many others, this body of work includes studies that assess a monitoring program for teacher attendance (Duflo, Hanna, and Ryan, 2007), ability grouping (Duflo, Dupas, and Kremer, 2008), textbooks (Glewwe, Kremer, and Moulin (2003), and teacher incentives (Muralidharan and Sundararaman, 2008). Compared to this body of work, this study is most closely aligned with Bannerjee, Duflo, Cole and Linden (2006) and He, Linden, MacLeod (2008) both of which assess changes in pedagogy. The former focused on math skills and providing remedial support for low performing children, and the latter assesses teaching strategies to deliver English language instruction to primary school children via a variety of delivery mechanisms.

The remainder of the paper is organized as follows. Section 2 provides an overview on the early childhood literacy literature and Section 3 provides a description of the *Shishuvachan* intervention. Section 4 provides a description of the research design including the experimental design, data collected, and statistical models. We describe the sample and assess the internal validity of the study in Section 5. And Section 6 provides estimates of the effects of the interventions. Finally, we conclude in Section 7.

II. Early Literacy

The importance of early reading abilities is well established in the literature. In a review of this research, Scarborough (2001) conducts a review of 61 research samples and shows a high correlation between kindergarten reading skills and later reading scores. Other studies such as Cunningham and Stanovich (1997) show that there are strong associations between first grade reading abilities and later reading skills as well as achievement in the 11th grade; Storch and Whitehust (2002) show that pre-literacy skills (e.g. knowing that writing goes from left to right and top to bottom on a page) is important for reading comprehension later on. Overall, there is a wealth of evidence showing a strong persistence in reading abilities starting from a young age.

Early strong readers tend to become strong readers later on and children with early reading disabilities tend to have reading disabilities even later into their academic career. In particular, vocabulary development may play a uniquely important role. Biemiller (2006) shows evidence that early vocabulary development may be important for reading comprehension later on. Despite this evidence, Biemiller and Slonim (2001) find that among preliterate children vocabularies can differ by several thousand root-meanings, a gap that may be difficult to close. So even though this literature has not demonstrated an experimentally validated causal relationship between early and late reading ability, there is strong suggestive evidence of a causal link between early and late reading. To the extent that later reading ability is important, so too is early reading.

This connection has strong implications for low-income children in particular. Because low-income children tend to have home and school environments less conducive to early reading development relative to those of higher income children, low-income children develop lower pre reading abilities which then have strong implications for later reading abilities. Capizanno et al (2000). Burchinal et al (2000) and Dickinson and Sprague (2001) show that the quality of early

childhood classrooms is important for disadvantaged children from infancy to the preschool years as these classroom settings may affect vocabulary and early literacy development. This research highlights the importance of targeting early reading development for lower-income students.

Given this problem, previous research also shows that early intervention can be effective at addressing early reading difficulties. Vellutino and Scanlon (2001) conduct a detailed study in which poor readers were identified and then randomly subjected to early one-on-one tutoring to address their reading weaknesses. The intervention proved effective, implying that rather than being a genetic disposition, reading skills can be developed. Even though one-on-one tutoring may not be feasible in low-income communities in developing countries, research shows that early intervention may be effective and important.

III. Background - The Pratham *Shishuvachan* Program

This study evaluates the Pratham *Shishuvachan* preschool program. *Shishuvachan* is a program conducted by the Pratham Mumbai Education Initiative, an educational NGO founded in 1994 in the slum communities of Mumbai. Its motto is "every child in school and learning well."[3] Over the years it has grown into a large organization with operations in 14 Indian states funded by the ICICI Bank of India, local governments, a large collection of corporate sponsors and international organizations, and local citizens. In addition to the Shishuvachan program, Pratham implements programs to mainstream out-of-school children into schools, programs

[3] http://www.pratham.org/aboutus/aboutus.php

teaching English to children using interactive electronic instruction booklets, and a variety of other educational programs.

The *Shishuvachan* program was originally designed to supplement *balwadi* (daycare) programs run by Pratham, the government school system, and other organizations. The program's main goal is to develop reading and comprehension skills among children aged four to five in preparation for primary school matriculation. It consists of two main components: the *Shishuvachan* classes and the *Bal vachanalaya* (child library). In the class, teachers use story books, flash cards for word and letter recognition, and *barakhadi* charts[4] to instruct children. The program aims to help children cognitively attach meaning to words and encourages children to take in whole words and ideas rather than overwhelming them with an abstract alphabet chart. The instructors engage in seven different activities: pre-reading, story-telling, story-reading, word recognition, letter recognition, *barakhadi* chart, and unfamiliar text reading.

The curriculum is precisely scheduled with specific activities conducted at specified periods in the program. Supervisors meet instructors twice a week, and four zonal heads meet supervisors once every ten days to ensure consistency in training and implementation. Both the class and child library are offered in the community and are accessible to all age-appropriate children. By participating in these two program components, children are expected to be able to read and comprehend simple unfamiliar stories after six months.

In practice, the *Shishuvachan* method may not have a uniform effect on students independent of the population of students or the method of delivery. Rather, a treatment can

[4]In Indian languages that use the Devanagri alphabet, each character designates a specific syllable. As an *abugida* script, the characters consist of a root consonant letter modified by diacritics indicating the different vowels. The chart is a standard illustration of the pattern of modifications made to each character to represent the universe of consonant-vowel matches. Learning this chart is the functional equivalent of learning the English alphabet.

interact differently with particular program structures and types of students to have larger or smaller effect on student outcomes (He, Linden, and MacLeod, 2008; Linden, 2008; Duflo, Dupas, and Kremer, 2009). Therefore, to test the general efficacy of these techniques, we evaluate the program in four different contexts over three years. First, we test the program on young children in their first year of the Indian public school system. Two interventions are compared to a control: an in school program and a community-based out-of-school program. In the second year, we estimate the effects of the techniques on students attending special classes designed to offer these techniques. Finally, in the third year, we estimate the effects of the techniques within the originally targeted population of pre-school children in established Pratham or government preschool classes.

IV. Research Design

To evaluate the *Shishuvachan* techniques in each of the contexts described in Section III, we conducted a three year study comprising three separate randomized controlled trials. In each trial, the basic strategy is to randomly assign students to either a treatment group that receives the treatment or a control group that does not. The random assignment should ensure that the groups are similar in all characteristics except for their receipt of the treatment, avoiding the common problem of self-selection into the treatment that must be considered in retrospective studies. Any difference in the children in each group after the implementation of the treatment can then be causally attributed to the program. In what follows, we first describe the general research design for each year, the data collected to evaluate the programs, and finally the statistical models used to assess the differences in student performance between the research groups.

A. Experimental Design

In the first year, 67 Mumbai municipal schools were randomly assigned to one of three research groups. Twenty-four were randomly assigned to receive the treatment on an out-of-school basis, 23 received the program on an in-school basis, and the remaining 20 schools constituted the control group. Enrollment and pre-school history was collected in September 2004. Schools were assigned to receive *Shishuvachan* methodology training for their instructors, to have a Shishuvachan program implemented on an out-of-school basis in the feeder community, or to receive no intervention at all. In all cases the Shishuvachan program was conducted for one hour a day.[5] After random assignment, 2679 students were assessed in January 2005 (pretest). The intervention was conducted over a period of six weeks from February to March. Beginning the third week of March, 1883 students took the post test. These figures are displayed in the first three columns of Table 1.

The intervention evaluated in the second year of the study involved testing the techniques in specially designed Shishuvachan classes, independent of any other educational infrastructure. Random assignment was done at the community level presenting unique econometric challenges. Because enrollment in the classes is voluntary and the control communities would not receive classes, it is impossible to define the sample to restrict attention to just those children who would participate in the classes if offered (as in we do in the first and third years). As a result, we use as our sample all students completing a baseline community census prior to randomization and methodologically, estimate the effect of the treatment on treated (Imbens and Angrist 1994).

[5] Indian schools operate for six days a week with the final day typically shorter than the first five.

In January of 2006, we conducted a baseline assessment of students in 138 slum communities in which Pratham operated preschools that had not received the *Shishuvachan* program. After this baseline assessment, the 4,757 children that were given the baseline assessment and who indicated an interest in the *Shishuvachan* classes were randomized by community into treatment and control groups, stratified by the percentage of children enrolled in the Pratham preschools. Seventy schools were selected for the treatment group and 68 for the control (columns four and five of Table 1). The *Shishuvachan* classes were started in February of 2006 and operated independently of existing preschools even in separate locations. As shown in the second panel of Table 1, only 29.2 percent of the students in the treatment communities took up the treatment.

In April of 2006, the post-program test (post-test) was administered to all children in both the treatment and control communities. Efforts were made to find as many of the children as possible who were also administered a pre-test including checking at their last known residence. Nonetheless, there was a relatively high level of attrition (748 students were lost at follow up), but as we will show in Section IV, the attrition patterns were the same in both treatment and control groups.

Finally, in the third year experiment, we evaluate the effects of implementing the *Shishuvachan* program on the children for which it was originally designed – preschool children in the Pratham (and some government-administered) preschool classes. Besides the context, the experiment also differed in the type if information collected – collecting information on parents reaching scores as well as those of the students enrolled in the classes.

In July 2006, prior to the pretest, surveyors met parents (or guardians) of three to five year old children in 235 communities at their homes and administered the same literacy test to establish the degree of literacy in participating households. This was done as a community census prior to starting the preschools in every community. The baseline test was conducted in August capturing information on 3,858 children attending the pre-school available in their community (columns six and seven of Table 1). One hundred sixteen communities were then selected for the treatment by a randomized selection process that was stratified by community level enrollment in the pre-school. After the randomization, the *Shishuvachan* techniques were introduced via specially trained instructors to all of the pre-schools. Surveyors conducted a post test in March 2007, surveying 2,907 of the original 3,858 students.

B. Description of the Data

Similar data was collected in all three years by a local group of assistants. To ensure objectivity, this group of assistants operated independently of the group implementing the program, having been recruited from Pratham activities in other parts of Mumbai. For all three years, three types of data were collected: the baseline assessment, demographic information, and the follow-up assessment. In 2007, data on the student's parent or guardian was also collected.

The baseline assessment involved giving the preschool aged children in the communities a short verbal test to determine their reading level, the standard test that Pratham uses in all of their early reading programs. Although the test was offered in a variety of languages, the vast majority of the children took a Hindi, Marathi, or Urdu test. The test was divided into four sections. The first section involved testing children in their knowledge of the alphabet. The

second section tested the children's ability to recognize words. The third section tested paragraphs and the fourth section tested stories of multiple paragraphs. The children's reading level was then calculated to either be nothing, letter, word, paragraph, or story depending on the highest section that the children could pass. Passing required that children be able to correctly identify the majority of the section. The follow-up assessment was the same as the baseline assessment.

Based on the results of this assessment, children are categorized into one of four levels: letter, words, paragraph, or story. In the first year of the project, only this categorization was recorded. However, in the subsequent two years, we recorded both the assigned level as well as the individual responses to each question asked of the student, providing additional variation in the children's scores. To account for this, we report scores for the last two years both in terms of the reading level of the child and in terms of the normalized score based on the number of correct answers provided by the children.

Demographic information was collected throughout each study year. This information included the student's gender, age, language of instruction, community of residence, and the type of pre-school the student attends. In 2007, additional data on the student's parent or guardian was also collected. In this data set, the parent or guardian was given the same assessment as the students. In addition, their educational attainment, relationship to the student, spoken language, and whether they had a student in the program were collected.

C. Analytic Models

In the first and third years of the study (2005 and 2007), the treatment was delivered through existing educational institutions with perfect compliance. This allows us to estimate the effect of the program for students attending those institutions by directly comparing the average performance of the treatment and control groups. Similarly, for the first year, we can directly compare the two treatment strategies by comparing the average performance in those groups. To do so, we use a simple difference estimator that takes the following form:

$$Y_{ij} = \beta_0 + \beta_1 Treat_j + \varepsilon_{ij} \qquad (1)$$

In this model, Y_{ij} is a student characteristic (such as gender, normalized pre-test scores, etc.) for student i in school j, $Treat_j$ is an indicator variable set to one if the respective institution j was assigned to treatment. In 2005, because there were three research groups, this model is estimated three times: once comparing the control group and in-school intervention group, once comparing the control group and out-of-school intervention group, and once comparing the in-school intervention and out-of-school intervention group.

We estimate the equation using ordinary least squares, and the estimator is used to both compare student characteristics at baseline and to compare the post-test results across the various research groups. An additional variant is used to control for observable student characteristics when comparing the follow-up test scores. The model is identical to the one presented in equation (1), but includes a vector of socio-demographic characteristics in addition to the treatment indicator.[6]

[6] To avoid dropping students if they were missing the full vector of socio-demographic information, we include an indicator variable for whether a given measure is missing and assign the respective control variable a value of zero. This information is provided in Table 3. For most measures, we are missing information on 1-2 percent of the sample, but for the second year experiment, we are missing language information on 34 percent of the sample.

Even if the randomization created initially similar treatment and control groups, differential attrition patterns could threaten internal validity. To estimate the possible effects of attrition, we compare the relative patterns of attrition in the treatment and control groups based on the information collected at baseline. To do so, we estimate the following equation using ordinary least squares:

$$Y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 Attrit_{ij} + \beta_3 Treat_j * Attrit_{ij} + \varepsilon_{ij} \quad (2)$$

In this specification, $Attrit_{ij}$ is an indicator set to one if the student did not give a post-test. The coefficient β_3 is a difference in differences estimate of the relative differences between attritors and non-attritors in the control and treatment groups.

In the second year of the study (2006), the assessment was done at the community level rather than at the school or pre-school level. Because not all students in the community attended the special *Shishuvachan* classes, the treatment estimates resulting from a simple difference estimate like that in equation (1) will likely under-estimate the treatment effect for students attending the classes. Since the goal of our study is to compare the estimates with those of the other years of the study which estimate treatment effects for students attending specified educational institutions, we follow (Imbens and Angrist 1994) and estimate the local average treatment effects on those students who attend the special classes. In addition to equations 1-3, we also estimate the following instrumental variable model:

$$Takeup_{ij} = \delta_0 + \delta_1 Treat_j + \mu_{ij} \quad (4A)$$

$$Y_{ij} = \beta_0 + \beta_1 Takeup_j + \varepsilon_{ij} \quad (4B)$$

In this model, $Takeup_{ij}$ is a dummy for whether individual *i* in community *j* actually took-up the treatment. Equation 4A is the first stage estimate where $Takeup_{ij}$ is run against the treatment dummy. Equation 4B is then the second stage estimate. As with equation (1), we also estimate a version of this controlling for demographic characteristics.

In all of the data we collect, students within the same community classroom share a number of experiences (same teacher, same classmates, same class resources, etc.) that cause their scores to be correlated. Without taking this into account, this correlation will cause us to overestimate the precision of the estimated treatment effects (Bertrand, Duflo, Mullainathan 2004).We correct for this by allowing the standard errors to be correlated at the level of randomization, the community or school level.

V. Sample Description and Internal Validity

A. Description of the Samples

The estimates in Tables 2 through 5 provide information on the academic abilities of the students in our sample in the first column of results for each experiment. Starting with Table 2, students in the first year sample score higher than students in the next two years which, given that they are older and have started a formal government school, seems reasonable. That said, the all of the children's reading levels are very low. Even in the first year sample, only 26.4 percent of students can do more than identify letters. The majority of the students can identify letters, but 18.2 percent of the students cannot even do that. In the later years, children perform even worse initially, with the vast majority having no reading skills.

Looking to the control averages in Table 3 provides information on the demographic characteristics of the sample. As expected, the characteristics of the sample of children in each year differ significantly. Students who attend the schools in the first year experiment generally do not attend a pre-school. While the majority of those in the second and third years do attend pre-school. Within these experiments, children in the third year are overwhelming more likely to attend a Pratham preschool (those that do not, attend a government preschool that was included in the study) while those children in the second year experiment are a bit more likely to attend a preschool run by an NGO other than Pratham, but equally likely to attend a government run preschool. In terms of language, students in the first year are primarily Marathi and Urdu speakers while those in the second and third years are equally likely to speak Marathi, Hindi, or Urdu. Finally, the distribution of gender is very similar across all of the samples. The point of the multiple experiments was to test the robustness of the intervention by estimating the effects on a wide variety of students – and as designed, the sample captures just this variety.

These differences are also evident in Table 5 which tabulates the characteristics of children who attend the different types of institutions in our study using the information available in the community census conducted in the second experiment. The first two columns show the difference between children attending and not attending the community-based classes in the second year of the study using only children from the treatment communities. Of the 28 percent of students who attended the classes from the treatment group, the children most highly represented were those with some but not significant reading skills – able to identify letters in the Pratham classification. Similarly, while the children attending the classes are equally distributed across the various preschool options, children in the community seem to be more heavily attending non-Pratham NGO preschools.

The second set of columns compares students who are attending a pre-school (either Pratham or a government pre-school) to those are that are not. This provides some information on the types of children that constitute the sample in the third year experiment in which the intervention is evaluated in existing pre-schools. Overall these students seem to be very similar. The aggregate average reading level is the same. Not surprisingly given their age, most the children either have no reading skills as measured by the assessment or are at the letter identification level. The only difference between the two groups seems to be that the children attending pre-school are slightly more likely to have no reading skills.

Finally, Table 4 contains the estimates of family members' reading skills. As Panel A shows, most of the data in this table (71.4 percent of the observations) came from one of the two parents. Some families refused to provide the information (16.7 percent), and the remainder of observations came from another adult family member who was not the child's parents. The parents' scores are presented in Panel B. The distribution of scores was bi-modal. Either parents were able to read at least a paragraph or they could not identify a letter. Within these groups, the parents were equally split. Finally, parents had, on average, about 4.5 years of schooling and about 1.2 children of age to participate in pre-school.

B. Baseline Comparison

This section considers threats to the internal validity of the study by comparing baseline characteristics. The internal validity of a randomized control study rests on the fact that the allocation of the treatment was independent of all student characteristics. While this statistical independence is imposed by construction in the randomization process, differences between

treatment and control group could arise as the samples are necessarily finite. If the differences are large enough, then the groups would differ outside of their treatment status preventing us from attributing causation of the difference to the treatment.

While we cannot check for comparability between the samples in each experiment along all dimensions, we can ask whether or not the groups differ significantly in the characteristics observed in our sample. The first evidence of comparability comes from Table 1. Looking to row three, the number of children per school or community is about the same. Tables 2 through 4 then provide more detail on the comparison between treatment and control groups. The general conclusion is that that overall the randomization seems to have worked well for all of the intended comparison groups with the out-of-school treatment in the first year experiment being the sole exception. And even this group differs from the other in only specific variable – child's home language. Given that this is the only significant difference, the results are what would be expected from random assignment, and we will be careful to control for these differences when comparing the out-of-school treatment group to the other two research groups in that year.

Examining the students' baseline performance on the assessment tool, we see a roughly similar distribution of students in Table 2. For each experiment, Table 2 contains the percentage of students in each reading category in Panel A (categories are omitted in year 2 and 3 because no students fell into those categories). To better gauge the magnitude of these differences, Panel B estimates a normalized version of this score in which students are assigned a grade of 0, 1, 2, 3 4 for being classified in the levels of "No Skills" to "Story Level" respectively. Results for the first, second, and third experiment are provided in columns one through five, six through eight, and nine through eleven respectively. In each case, we provide an estimate of the control mean and the difference between treatment and control group, as well as the difference in research

groups for the first year of the study. The point of these comparisons is to determine whether there are differences in the respective research groups large enough to indicate that one group is likely to score much higher than the other on the post-test, absent the treatment. To provide a gauge of these magnitudes, the last column for each year contains the correlation between the normalized post-test scores and the observable characteristics using students from the control group.[7]

Comparing the treatment and control groups, the groups are fairly well balanced. The differences in the second and third year experiments are very small with a normalized difference of 0.06 standard deviations and 0.05 standard deviations respectively. For each category, the percentage difference is less than 2 percentage points. None of these differences are statistically significant. Some of the differences in the first year are statistically significant, but they are too small to pose a threat to internal validity. For example, consider the 10.7 percentage point difference in fraction of students who can read letters between the out-of-school and in-school intervention groups. The marginal correlation between testing at this initial level, and student's the normalized follow-up score is only 0.425 standard deviations. A 10.7 percent difference at baseline is thus consistent with only a 0.045 standard deviations difference in the mean student score on the follow-up assessment.

Table 3 reports the same comparison using the demographic controls. For each of the groups except the out-of-school treatment in the first year experiment, the groups are again sufficiently similar that we anticipate no large differences at follow-up for reasons other than the receipt of the treatment. Only two differences are large enough to be statistically significant in

[7] The estimates are from a regression that includes all of the covariates in Tables 2 through 4. The normalized post-test score used is the measure taken from the classification of students with respect to their reading abilities similar to that used in Panel B of Table 2.

the first year. First is the 16 percent difference in the probability that the out-of-school treatment children attended a preschool run by an NGO other than Pratham. Second, children in the out-of-school treatment are 30 percentage points less likely to have been Marathi speaking. Compared to the differences observed in Table 2, these differences are larger – implying a net difference in normalized test scores of about 0.18 standard deviations at follow-up. However, it is important to note that this difference will work against the out-of-school treatment because Marathi students score higher on the follow-up test than other students. As a result, however, we will be sure to focus on the estimates that control for these demographic characteristics.

The baseline differences in the second and third year experiments are similar in magnitude to the differences between the in-school program and the control group in the first year experiments. While a few of the differences are large enough to be statistically significant, none are large enough to generate significant differences in the research group at follow-up based on the correlations presented for each year. The largest difference, for example, is the language difference in the second year. In this case, the treatment group is 11.6 percentage points more likely to be Marathi speaking. However, using the differences in the other language categories and the provided correlation with the follow-up test score, the implied difference at follow-up is only 0.031 standard deviations.

Table 4 compares parents' scores and family characteristics in the third experiment. As before the differences are generally very small. Interestingly, as shown in column three, children of parents who scored at the paragraph level or above have a reading score about 0.358 standard deviations higher than those who families are less well educated. However, the distribution of parents' scores across the two groups is identical. The treatment group families have about 0.07

more children in the eligible age range, but again while the difference is statistically significant, the magnitude of the effect is small.

C. Attrition Patterns

Even if the various research groups are comparable at baseline, differences in the types of students that attrit from the sample over the course of the experiment could violate the internal validity of the study if those differences are correlated with the receipt of the treatment. (For example, if poorly performing students were less likely to exit the sample in the treatment group due to the intervention). To check for this, we compare the attriting students in each research group to the non-attriting students and then compare the difference in characteristics across the research groups in each experiment using equation (2). These results are presented in Table 6 for the test score differences. In results not presented, we calculate the difference for each of the variables in Tables 2, 3, and 4. The results show no differential attrition patterns.

The first thing to note is that in Panel A, the fraction of attriting students is very similar across the research groups. The difference is largest is the third year and is statistically significant at 5.2 percentage points, but the magnitude of the difference is small. Comparing students across the reading categories, only two differences are large enough to be statistically significant ("No Skills" in year one in-school versus control and "Read Letters" in year three), but again , the differences are small. The same is true for the normalized scores in Panel C.

VI. Differences at Follow-Up

As long as the assignment of the treatment is allocated independently of the characteristics of each student, we can estimate the causal effect of the *Shishuvachan* interventions by directly comparing the average scores in the treatment and control groups. In what follows, we first estimate the overall effects calculating the net changes in reading skills resulting from the program as well was the normalized change in reading scores. Then we use the demographic information and information on parents' reading abilities to assess the distribution of gains within subsets of students.

Tables 7, 8, and 9 present the main finding that children participating in the program display greater gains than their counterparts in the control groups for the experiments in years one, two, and three respectively. In all three tables, Panel A presents the results for the full sample regardless of the individual's baseline score. Panel B restricts the sample to those individuals who scored at the "Nothing" level; Panel C is restricted to individuals whose highest level attained was "Letter Identification;" and so forth. In each panel, we omit the students with "No Skills" since this category is simply the complement of those presented. Sample sizes for each group are reported in the last row of each panel. The columns show the usual differences in treatment groups over control groups, but are reported separately for two models, one without controls for baseline characteristics and one with these controls.

The results for the first year experiment are presented in Table 7. Starting with Panel A, both programs seem to generate positive gains at all levels of student performance. As expected, the coefficients in the model with controls and without controls are generally similar for the in-school versus control specification, but are generally higher for comparisons involving the out-of-school intervention due to the imbalance in this group noted in Section IV. However, within the model that controls for baseline characteristics, both programs increase the percentage of

students in every category and the point estimates are larger for the out-of-school intervention in each category except Story Level. The differences are largest in the Letter and Word categories and the difference is statistically significant both as well as for the Story level, showing that the out-of-school intervention seems to be more productive than the in-school intervention.

To better assess the gains achieved by each type of student, Panels B, C, and D show the final reading categories divided by the student's starting category. For students with no baseline reading skills in Panel B, the model with the control variables indicates that only the out-of-school intervention increased reading scores for students at this level. While the point estimates for the in-school intervention are all positive, they are all statistically insignificant and less than ten percentage points. The out-of-school intervention increases 35 percent of these students to the letter level and 30 percent to the word level both of which are statistically significant at the one percent level. Both represent statistically significant gains over the in-school intervention. For students who can identify letters or words at baseline, both programs generate statistically significant gains. The out-of-school intervention outperforms the in-school intervention for the Letter level students, but both programs seem to benefit the fifteen percent of students at Word level at baseline similarly.

Table 8 contains the estimated treatment effects for the *Shishuvachan* community classes evaluated in the second year experiment. As explained in Section III, the sample for this experiment was all of the children in a given community while not all children in those communities actually took up the intervention. This result is that the average treatment effect estimated at the community level is not comparable to the average treatment effects presented in Tables 6 and 9 because these latter tables estimate treatment effects only on children who self selected into the respective institutions. We thus present two different estimates for this sample.

We present the overall community-level treatment effect estimated with equation (1) in columns two and five and present the treatment effect for students attending the special *Shishuvachan* classes (estimated with equation (3)) in columns three and six. The latter estimates more closely conform to the estimates generated from the other experiments.

Compared to the effects of the program in the public schools, the gains in the special *Shishuvachan* classes are limited but still positive. The overall estimates presented in Panel B show that while the program generates more children in each reading category, the effect is only large enough to be statistically significant at the Letter level. The classes increase the number of children at this level by 9.9 percentage points (with baseline controls) a difference that is statistically significant at the five percent level. For students actually attending the class, the difference is 35.9 percentage points. This is a large impact, but unlike the school-based interventions, the effects seem to be limited to this level.

Panels B through D provide more detail on the distribution of gains. The benefits are isolated to just those students who had no reading skills at baseline and while the program seems to have enabled them to read letters, the gains are limited to that skill level. Sixteen percent of all children in the community who had no reading skills could identify letters as a result of the program. For students who actually attended the classes, this represented 47 percent of students. Both differences are statistically significant at the one percent level. For students at the Letter Level initially, the program seems to have made no difference in their reading levels. The same is true for the very few students at Word level, although the small number of students makes it difficult to detect an effect for these students.

Table 9 presents the results for the pre-school based program. Like the school based programs, the impact of the intervention in this setting is both large and robust. Panel A again presents the overall effects for all children. Compared to the control group, 22.2 percent more students in the treatment preschools are able to identify letters, 19.6 percent are able to identify words, 9.5 percent can read at the Paragraph level, and 3.3 percent more (for a total of 5.2 percent) can read at the Story level. All of these differences are statistically significant at the one percent level. Given that the initial reading abilities and ages of this sample are similar to those in the *Shishuvachan* stand-alone-class experiment, these results confirm that in the right setting these teaching methods can have a very robust impact on student performance.

Dividing the sample by initial performance in Panels B, C, and D, the results are concentrated in the 93 percent of students who could not read even letters at baseline. The effects for these students are presented in Panel B. The results are slightly larger, but similar to the effects in Panel A with students much more likely to be categorized into each of the presented reading categories. Panels C and D present the results for the few students with higher skill levels at baseline. The results are generally mixed, but due to the small number of students, the standard errors are fairly large. Only one difference is statistically significant and this shows 28.4 percent decline amongst students who are classified at letter level for students who are at word level at baseline. This may suggest that the intervention does a good job in helping these high performing students to retain their reading skills, but given the lack of consistent results, there is no other support for this result among these higher performing students.

While the reading level is a useful absolute characterization of students' skill levels, it is difficult to compare these results to those of other interventions. Table 10 presents the same results presented earlier, but as a normalized scored relative to the control group distribution for

the respective experiment. As explained in Section IV, we have different levels of detail on the students' follow-up test scores in the first versus second and third years. In the first year, we only have the students' overall score which is a ranking from 0 to 4 that indicates whether the student has no reading abilities (0), is at Letter level (1), etc. For the second and third year we have the number of correct answers to the individual sections of each follow-up assessment. As a result, we present two estimates of the treatment effect in Panel A. For the first year we report the treatment effect in terms of the normalized total score and for years two and three we report both the normalized total score (row one) as well as a normalization of the total number of individual questions answered correctly (row two). For the second two years, we also present the normalized score for the individual score on each section of the follow-up test.

The results are consistent with those in Tables 7, 8, and 9. In the first year experiment, the in-school intervention and the out-of-school intervention are both very effective strategies with the in-school intervention generating a gain of 0.26 standard deviations and the out-of-school intervention improving scores by 0.55 standard deviations. Both are statistically significant at the one percent level – and the out-of-school intervention is clearly higher performing than the in-school intervention with the difference of 0.24 being statistically significant at the five percent level.

The community based intervention generates an effect of 0.12 standard deviations for the entire community and relative to the other interventions generates a difference for participating students of 0.44 to 0.53 standard deviation depending on the normalization. The overall change in test scores is not statistically significant, but the point estimate is similar in magnitude to those experienced by the students participating in the other educational institutions. As indicated in the previous tables, the treatment effect was concentrated amongst the lowest competencies. The

effect on letter identification was 0.77 standard deviations and is statistically significant at the five percent level.

The intervention seems to have proved most effective in the setting for which it was originally designed – pre-schools. The treatment effect in these institutions is 0.7 standard deviations overall, an effect that is larger than in either of the previous institutions. Consistent with the previous results, the program also generates large changes in the students' reading levels at all competency levels with the largest effects measured in reading words.

Finally, in Table 11, we disaggregate the effects of the pre-school intervention by parents' reading abilities. All of these estimates are presented using the normalize score variable presented in row two of Table 10. Young children have limited opportunities for educational inputs outside of their immediate families – resulting in the correlation between children's reading skills and their parents' observed in Table 4. These disparities between children might also be correlated with the degree to which the children respond to the intervention. For reference, the first row shows the overall average treatment effect for this experiment, 0.74 standard deviations.

The results indicate that rather than being correlated with large differences in the effectiveness of the treatment, all children seem to benefit from the program in a similar manner. Children with parents with no reading skills benefit by 0.84 standard deviations while children with parents who can read at the Story Level experience a gain of 0.65 standard deviations. Both effects are very large, and children whose parents have lower reading ability benefit by 0.19 standard deviations more, but the difference is not statistically significant. A similar result emerges if we estimate the relationship using parents' education levels. Children with

uneducated parents experience a larger treatment effect – 0.82 standard deviations. But even children with parents educated between 6 and 10 years still experience a gain of 0.693 standard deviations. Children of parents with the highest educational attainment experience the smallest treatment effect, but there are so few children in this category that the standard errors are over twice as large as those for the point estimates for the other categories. The results confirm that all children benefit significantly from the teaching methods regardless of their parents reading abilities and educational backgrounds.

VI. Conclusion

The results of this experiment suggest that the Pratham *Shishuvachan* early literacy skills development curriculum is a viable strategy for improving the reading skills of pre-school and first grade children in India. In the three experiments used for the evaluation, the program proves robust to several delivery mechanisms, delivering gains of 0.12 to 0.70 standard deviations over the respective control groups.

Comparing the different implementation strategies, we see the greatest gains when the program is a supplement to existing language training and among children with the lowest initial performance or whose parents are least able to supplement their classroom studies. The in-school intervention of year one demonstrated a gain of 0.26 standard deviations compared to the 0.55 standard deviation gain of the out-of-school intervention. Providing Shishuvachan as a complement to school curricula rather than as a substitute has a significantly more positive effect on student scores. The community based intervention of year two generated results show an insignificant increase in normalized score (0.12 standard deviations for the entire community,

0.44 to 0.53 standard deviations for participating students). In all years the strongest gains are by students who begin at the lowest baseline reading ability.

VII. Bibliography

Angrist, Joshua D.; Imbens, Guido W.; Rubin, Donald B. “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, June 1996.

Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2004) "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, , *119*(1), pp. 249-75.

Biemiller, A. (2006). Vocabulary Development and Instruction: A Prerequisite for School Learning. In S.B. Neuman and D.K. Dickinson (Eds.), *Handbook of Early Literacy Research (Volume 2)* (pp. 41-51). New York, NY: Guilford Press.

Biemiller, A., and Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *94*, 465-473.

Burchinal, M.R., Roberts, J.E., Riggins, Jr., R., Zeisel, S.A., Neebe, E., and Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*, 339-357.

Capizzano, J., Adams, G., and Sonenstein, F. (2000). Child care arrangements for children under five: Variation across states. Washington, DC: The Urban Institute. (Book)

Cunningham, A.E. and Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934-945.

Dickinson, D.K. and Sprague, K.E. (2001). The nature and impact of early childhood care environments on the language and early literacy development of children from low-income families. In S.B. Neuman and D.K. Dickinson (Eds.), *Handbook of Early Literacy Research (Volume 1)* (pp. 263-280). New York, NY: Guilford Press.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, (2008). “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.” Poverty Action Lab, Cambridge, MA: Working Paper.

He, Fang, Leigh Linden, and Margaret MacLeod, “How to Teach English in India: Testing the Relative Productivity of Instruction Methods within the Pratham English Language Education Program” Columbia University: Working Paper, 2008

Hanushek, Eric, and Ludger Wooesman. (2007) “The Role of Education Quality for Economic Growth.” World Bank Policy Research Working Paper No. 4122

Imbens, Guido W & Angrist, Joshua D, (1994). "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, Econometric Society, vol. 62(2), pages 467-75, March

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

Storch, S.A. and Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from longitudinal structural model. *Developmental Psychology, 38(6),* 934-947.

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

Table 1: Description of Samples, All Years

	Year 1: 2005			Year 2: 2006		Year 3: 2007	
Sample Decription	**Control**	**School**	**Community**	**Control**	**Treatment**	**Control**	**Treatment**
Number of Classes/Communities	20	23	24	70	70	119	116
Number of Students	636	746	707	2458	2517	1863	1995
Number Students Treated	0	746	707	0	734	0	1995
Number of Students Untreated	636	0	0	2458	1783	1863	0

Table 2: Baseline Test Scores, All Three Years

	Year 1: 2005-06					Year 2: 2006-07			Year 3: 2007-08		
Characteristic	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**	**Post-Test Correlation**	**Control Mean**	**Treatment - Control**	**Post-Test Correlation**	**Control Mean**	**Treatment - Control**	**Post-Test Correlation**
Panel A: Skill Level Category											
No Reading Skills	0.182	-0.052	0.020	0.072*		0.605	-0.01		0.932	-0.009	
		(0.042)	(0.050)	(0.040)			(0.044)			(0.024)	
Read Letters	0.368	0.066	-0.040	-0.107**	0.425***	0.369	-0.009	0.404***	0.057	0.004	1.322***
		(0.045)	(0.052)	(0.053)	(0.102)		(0.039)	(0.105)		(0.019)	(0.275)
Read Words	0.147	0.056*	0.018	-0.038	0.968***	0.025	0.016	1.415***	0.012	0.004	2.223***
		(0.033)	(0.039)	(0.042)	(0.125)		(0.014)	(0.370)		(0.007)	(0.758)
Read Sentences	0.088	-0.007	-0.048*	-0.041*	1.671***	0.002	0.004	0.804***			
		(0.034)	(0.028)	(0.024)	(0.163)		(0.004)	(0.214)			
Read Paragraph	0.029	0.002	-0.012	-0.014	1.887***				0.000	0.001	0.000***
		(0.017)	(0.017)	(0.010)	(0.163)					(0.001)	(0.000)
Panel B: Total Score											
Normalized Score	< 0.001	0.089	-0.151	-0.240*	0.621***	0	0.06	0.320***	< 0.001	0.050	0.379***
		(0.156)	(0.166)	(0.134)	(0.050)		(0.100)	(0.055)		(0.096)	(0.095)

Table 3: Baseline Characteristics, All Three Years

	Year 1: 2005-06					Year 2: 2006-07			Year 3: 2007-08		
Characteristic	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**	**Post-Test Correlation**	**Control Mean**	**Treatment - Control**	**Post-Test Correlation**	**Control Mean**	**Treatment - Control**	**Post-Test Correlation**
Panel A: Pre-School Type Attended											
Pratham	0.121	-0.034	-0.037	-0.003	0.061	0.213	-0.046	0.206**	0.885	0.070*	-0.217
		(0.024)	(0.022)	(0.018)	(0.109)		(0.039)	(0.103)		(0.036)	(0.241)
Other Non-Government	0.343	-0.124	0.040	0.164*	0.152	0.367	0.026	0.352**	0.035	-0.002	0.029
		(0.082)	(0.105)	(0.093)	(0.095)		(0.041)	(0.144)		(0.019)	(0.364)
Government						0.197	-0.004	0.161	0.126	-0.055	-0.134
							(0.042)	(0.110)		(0.042)	(0.236)
Public School						0.091	-0.038*	0.522***			
							(0.023)	(0.190)			
None	0.536	0.158*	-0.003	-0.161		0.127	-0.025		0.018	-0.017	
		(0.092)	(0.114)	(0.102)			(0.025)			(0.013)	
No Information						0.013	0.094***	0.149			
							(0.020)	(0.175)			
Panel B: Language Spoken at Home											
Marathi	0.491	-0.145	-0.297**	-0.152		0.163	0.116**		0.313	-0.062	
		(0.162)	(0.143)	(0.131)			(0.049)			(0.052)	
Hindi	0.083	0.158	0.187	0.029	-0.708***	0.171	-0.005	-0.041	0.307	-0.014	-0.058
		(0.108)	(0.114)	(0.135)	(0.086)		(0.043)	(0.111)		(0.052)	(0.091)
Urdu	0.426	-0.013	0.110	0.123	-0.430***	0.215	-0.023	-0.054	0.300	0.050	0.016
		(0.171)	(0.165)	(0.158)	(0.069)		(0.050)	(0.139)		(0.061)	(0.128)
Gujarati						0.001	-0.001	-0.455***	0.003	0.003	0.655
							(0.001)	(0.102)		(0.003)	(0.641)
English						0.106	-0.036	0.057	0.000	0.000***	0.000***
							(0.025)	(0.129)		(0.000)	(0.000)
Southern Language						0.004	-0.004	0.873***	0.029	-0.019*	-0.05
							(0.004)	(0.143)		(0.011)	(0.189)
No Information						0.34	-0.053	-0.662***	0.001	0.019	0.000***
							(0.040)	(0.126)		(0.013)	(0.000)
Panel C: Child's Gender											
Male	0.521	0.030	-0.029	-0.059	-0.039	0.485	0.014	0.009	0.443	0.041**	0.047
		(0.036)	(0.036)	(0.037)	(0.070)		(0.021)	(0.049)		(0.017)	(0.060)

Omitted language for all three years is Marathi, the official state language. In year 2005 the variable records the medium of program instruction. IN 2006 & 2007 it refers to the student's mother tongue as instructors saught to speak to students in multiple languages.

Table 4: Family Characteristics, Year 3 (2007-08)

Characteristic	Year 3: 2007-08 Control Mean	Treatment - Control	Post-Test Correlation
Panel A: Parents Reading Ability			
Parent Normalized Score	0.000	0.040	-0.111
		(0.072)	(0.099)
Parent Reads less than a Paragraph	0.530	-0.014	omitted
		(0.036)	
Parent Reads Paragraph	0.470	0.014	0.358**
		(0.036)	(0.181)
Panel B: Parent Taking Test			
Mother Took Test	0.614	0.014	-0.031
		(0.033)	(0.081)
Father Took Test	0.100	-0.010	-0.043
		(0.020)	(0.105)
Parental Score Missing	0.167	-0.006	0.138
		(0.026)	(0.131)
Panel C: Other Family Information			
Parents' Years of Education	4.505	-0.138	0.012
		(0.288)	(0.011)
No Parental Education Information	0.264	0.031	-0.001
		(0.034)	(0.080)
Number of Children Aged 3 to 5	1.192	0.073***	0.085
		(0.027)	(0.054)

Table 5: Relative Characteristics of Participating Students

	Treatment Children Attending Program			All Chidren Attending a Pre-School		
Characteristic	**Non-Attending**	**Attending**	**Difference**	**Non-Attending**	**Attending**	**Difference**
Panel A: Participation Rates						
Fraction of All Children	0.622	0.278		0.089	0.189	0.102***
						(0.025)
Number of Children	1711	659		2064	1594	
Panel B: Baseline Reading Levels						
Reading Level	0.531	0.264	-0.267***	0.48	0.37	-0.11***
			(0.058)			(0.039)
No Reading Skills	0.535	0.75	0.215***	0.556	0.658	0.102***
			(0.045)			(0.035)
Identify Letters	0.406	0.238	-0.168***	0.409	0.317	-0.091***
			(0.038)			(0.034)
Read Words	0.052	0.011	-0.041***	0.034	0.021	-0.013
			(0.014)			(0.011)
Read Paragraphs	0.007	0.002	-0.005	0	0.003	0.003
			(0.005)			(0.002)
Panel C: Student Characteristics						
Male	0.505	0.483	-0.023	0.517	0.448	-0.071***
			(0.026)			(0.018)
Pratham Pre-School	0.124	0.278	0.153***			
			(0.036)			
Other Non-Govt Pre-S	0.461	0.219	-0.242***			
			(0.035)			
Government Pre-Scho	0.198	0.181	-0.017			
			(0.046)			
No Pre-School	0.117	0.061	-0.056**			
			(0.026)			
In School	0.06	0.035	-0.025			
			(0.017)			
No Pre-School	0.049	0.258	0.209***			
			(0.038)			

Table 6: Relative Characteristics of Attritors and Non-Attritors, All Years

	Year 1: 2005-06				Year 2: 2006-07		Year 3: 2007-08	
Characteristic	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**	**Control Mean**	**Treatment - Control**	**Control Mean**	**Treatment - Control**
Panel A: Attrition Rates								
Number Attriting	66	37	61	24	602	146	288	124
Percentage Attriting	9.2%	2.0%	4.9%	2.9%	33.3%	-1.7%	18.9%	5.2%**
		(0.028)	(0.040)	(0.038)		(0.035)		(0.025)
Panel B: Skill Level Category								
No Reading Skills	0.170***	-0.163*	-0.08	0.083	0	0.041	0.055***	-0.037
	(0.051)	(0.098)	(0.121)	(0.095)	(0.024)	(0.050)	(0.015)	(0.024)
Read Letters	0.034	0.057	-0.023	-0.08	-0.01	-0.035	-0.044***	0.034*
	(0.063)	(0.077)	(0.074)	(0.071)	(0.024)	(0.044)	(0.014)	(0.019)
Read Words	0.027	0.072	0.073	0.001	0.01	-0.001	-0.011*	0.004
	(0.047)	(0.090)	(0.099)	(0.069)	(0.008)	(0.016)	(0.007)	(0.011)
Read Sentences	-0.054	0.028	0.057	0.029	0	-0.006		
	(0.038)	(0.043)	(0.053)	(0.041)	(0.002)	(0.004)		
Read Paragraph	-0.034	-0.001	0.031	0.032**			0.000***	-0.001
	(0.022)	(0.020)	(0.020)	(0.013)			(0.000)	(0.001)
Panel C: Total Score								
Normalized Score	-0.382***	0.272	0.368	-0.096	0.01	-0.054	-0.066***	-0.029
	(0.129)	(0.218)	(0.266)	(0.197)	(0.028)	(0.061)	(0.019)	(0.019)

Table 7: Student Skill Level at Follow-Up, Year 1 (2005-06)

	No Controls for Baseline Characteristics				Controlling for Baseline Characteristics			
Characteristic	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**
Panel A: All Students								
Identify Letters	0.845	0.017	0.063	0.046	0.845	0	0.115***	0.066*
		(0.042)	(0.043)	(0.035)		(0.033)	(0.041)	(0.039)
Read Words	0.505	0.082	0.159**	0.077	0.505	0.113***	0.282***	0.139**
		(0.084)	(0.077)	(0.070)		(0.044)	(0.057)	(0.058)
Read Paragraph	0.202	0.098*	0.082	-0.016	0.202	0.114***	0.155***	0.039
		(0.053)	(0.051)	(0.049)		(0.036)	(0.036)	(0.039)
Comprehend Story	0.075	0.049	0.032	-0.017	0.075	0.064**	0.063***	0.021
		(0.034)	(0.028)	(0.031)		(0.032)	(0.019)	(0.025)
Sample Size	560	1267	1176	1323	560	1267	1176	1323
Panel B: No Reading Skills at Baseline								
Identify Letters	0.453	0.104	0.393***	0.289***	0.453	0.093	0.346***	0.244**
		(0.097)	(0.093)	(0.088)		(0.102)	(0.104)	(0.098)
Read Words	0.128	0.089	0.384***	0.295***	0.128	0.079	0.295***	0.249***
		(0.081)	(0.098)	(0.099)		(0.071)	(0.071)	(0.091)
Read Paragraph	0.009	0.01	0.084***	0.074**	0.009	0.016	0.043**	0.039
		(0.014)	(0.030)	(0.031)		(0.016)	(0.021)	(0.025)
Comprehend Story	0.009	0.001	0.004	0.003	0.009	0.005	0.003	0.003
		(0.012)	(0.011)	(0.012)		(0.013)	(0.012)	(0.017)
Sample Size	117	223	179	268	117	223	179	268
Panel C: Letter Identification at Baseline								
Identify Letters	0.933	-0.06	-0.016	0.044	0.933	-0.079**	0.002	0.048
		(0.037)	(0.035)	(0.044)		(0.036)	(0.034)	(0.049)
Read Words	0.431	0.09	0.200**	0.11	0.431	0.099*	0.263***	0.153**
		(0.099)	(0.100)	(0.102)		(0.057)	(0.072)	(0.066)
Read Paragraph	0.071	0.088**	0.150***	0.062	0.071	0.098***	0.161***	0.072*
		(0.039)	(0.053)	(0.058)		(0.034)	(0.043)	(0.043)
Comprehend Story	0.004	0.028**	0.065***	0.037*	0.004	0.039***	0.060***	0.041***
		(0.012)	(0.017)	(0.020)		(0.013)	(0.014)	(0.016)
Sample Size	255	606	532	624	255	606	532	624
Panel D: Read Words at Baseline								
Identify Letters	0.96	-0.005	-0.015	-0.01	0.96	0.036	0.025	-0.007
		(0.031)	(0.039)	(0.040)		(0.024)	(0.021)	(0.026)
Read Words	0.772	-0.008	0.057	0.065	0.772	0.185***	0.231***	0.057
		(0.111)	(0.092)	(0.104)		(0.062)	(0.057)	(0.066)
Read Paragraph	0.327	0.151	0.146	-0.005	0.327	0.311***	0.267***	-0.02
		(0.114)	(0.108)	(0.113)		(0.085)	(0.084)	(0.071)
Comprehend Story	0.099	0.092	0.056	-0.036	0.099	0.213**	0.07	-0.035
		(0.072)	(0.065)	(0.070)		(0.091)	(0.055)	(0.058)
Sample Size	101	258	230	286	101	258	230	286

Table 8: Student Skill Level at Follow-Up, Year 2 (2006-07)

	No Controls for Baseline Characteristics			Controlling for Baseline Characteristics		
Characteristic	**Control Mean**	**Treatment - Control**	**IV Treat-Cont**	**Control Mean**	**Treatment - Control**	**IV Treat-Cont**
Panel A: All Students						
Identify Letters	0.42	0.149***	0.452**	0.42	0.098**	0.354**
		(0.054)	(0.176)		(0.047)	(0.168)
Read Words	0.141	0.039	0.118	0.141	0.019	0.067
		(0.040)	(0.124)		(0.038)	(0.136)
Read Paragraph	0.035	0.007	0.021	0.035	0.005	0.018
		(0.024)	(0.073)		(0.025)	(0.089)
Comprehend Story	0.019	0.011	0.034	0.019	0.008	0.03
		(0.017)	(0.052)		(0.017)	(0.060)
Sample Size	1,206	2,828	2,828	1,206	2,824	2,824
Panel B: No Reading Skills at Baseline						
Identify Letters	0.281	0.212***	0.505***	0.281	0.162***	0.470***
		(0.057)	(0.147)		(0.056)	(0.157)
Read Words	0.062	0.046*	0.111	0.062	0.037	0.106
		(0.028)	(0.068)		(0.032)	(0.090)
Read Paragraph	0.016	-0.004	-0.009	0.016	-0.005	-0.015
		(0.009)	(0.021)		(0.010)	(0.030)
Comprehend Story	0.005	0.004	0.01	0.005	0.003	0.01
		(0.006)	(0.015)		(0.006)	(0.019)
Sample Size	729	1,672	1,672	729	1,669	1,669
Panel C: Letter Identification at Baseline						
Identify Letters	0.619	0.043	0.192	0.619	-0.002	-0.011
		(0.059)	(0.273)		(0.056)	(0.318)
Read Words	0.245	0.017	0.079	0.245	0.002	0.01
		(0.061)	(0.277)		(0.060)	(0.338)
Read Paragraph	0.051	0.028	0.127	0.051	0.03	0.168
		(0.048)	(0.225)		(0.049)	(0.285)
Comprehend Story	0.033	0.023	0.103	0.033	0.022	0.125
		(0.036)	(0.167)		(0.035)	(0.207)
Sample Size	449	1,055	1,055	449	1,054	1,054
Panel D: Read Words at Baseline						
Identify Letters	0.846	-0.059	-0.723	0.846	0.018	0.443
		(0.113)	(1.455)		(0.094)	(2.470)
Read Words	0.538	-0.096	-1.169	0.538	-0.119	-2.933
		(0.148)	(1.835)		(0.182)	(6.357)
Read Paragraph	0.269	-0.138	-1.685	0.269	-0.121	-2.983
		(0.110)	(1.508)		(0.109)	(5.576)
Comprehend Story	0.154	-0.055	-0.677	0.154	-0.059	-1.446
		(0.096)	(1.174)		(0.080)	(3.167)
Sample Size	26	87	87	26	87	87

Table 9: Student Skill Level at Follow-Up, Year 3 (2007-08)

	No Controls for Baseline Characteristics		Controlling for Baseline Characteristics	
Characteristic	**Control Mean**	**Treatment - Control**	**Control Mean**	**Treatment - Control**
Panel A: All Students				
Identify Letters	0.243	0.204***	0.243	0.221***
		(0.032)		(0.030)
Read Words	0.068	0.199***	0.068	0.196***
		(0.020)		(0.021)
Read Paragraph	0.031	0.114***	0.031	0.095***
		(0.016)		(0.016)
Comprehend Story	0.019	0.050***	0.019	0.033***
		(0.013)		(0.011)
Sample Size	1,768	3,665	1,768	2,594
Panel B: No Reading Skills at Baseline				
Identify Letters	0.207	0.213***	0.207	0.232***
		(0.032)		(0.031)
Read Words	0.044	0.204***	0.044	0.200***
		(0.018)		(0.020)
Read Paragraph	0.019	0.115***	0.019	0.102***
		(0.015)		(0.014)
Comprehend Story	0.009	0.051***	0.009	0.041***
		(0.011)		(0.009)
Sample Size	1,647	3,398	1,647	2,421
Panel C: Letter Identification at Baseline				
Identify Letters	0.72	0.026	0.72	0.125
		(0.083)		(0.119)
Read Words	0.38	0.085	0.38	0.108
		(0.103)		(0.123)
Read Paragraph	0.15	0.052	0.15	-0.078
		(0.076)		(0.088)
Comprehend Story	0.11	-0.005	0.11	-0.097
		(0.061)		(0.077)
Sample Size	100	214	100	138
Panel D: Read Words at Baseline				
Identify Letters	0.762	0.038	0.762	-0.284***
		(0.200)		(0.089)
Read Words	0.524	0.11	0.524	0.171
		(0.233)		(0.180)
Read Paragraph	0.381	0.119	0.381	-0.029
		(0.241)		(0.243)
Comprehend Story	0.381	0.052	0.381	-0.029
		(0.247)		(0.243)
Sample Size	21	51	21	34

Table 10: Normalized Reading Scores at Follow-Up, All Years

	Year 1: 2005-06				Year 2: 2006-07			Year 3: 2007-08	
Test Sections	**Control Mean**	**School - Control**	**Community - Control**	**Community - School**	**Control Mean**	**Treatment - Control**	**IV Treat-Cont**	**Control Mean**	**Treatment - Control**
Panel A: Aggregate Scores									
Normalized Reading Level	0	0.261***	0.550***	0.237**	0	0.123	0.443	0	0.701***
		(0.092)	(0.109)	(0.116)		(0.123)	(0.447)		(0.082)
Normalized Total Score					0	0.146	0.528	0	0.695***
						(0.114)	(0.412)		(0.085)
Panel B: Normalized Score by Section									
Identify Letters					0	0.210**	0.756**	0	0.496***
						(0.093)	(0.329)		(0.075)
Read Word					0	0.031	0.111	0	0.814***
						(0.115)	(0.414)		(0.087)
Read Paragraph					0	-0.009	-0.033	0	0.605***
						(0.140)	(0.504)		(0.095)
Comprehend Story					0	0.031	0.113	0	0.304***
						(0.144)	(0.521)		(0.094)

Table 11: Treatment Effect Estimated by Parental Characteristics, Year 3

Test Sections	Control Mean	Treatment - Control
All Sections, Normalized Level Score	0	0.740***
		(0.078)
Panel A: Parental Reading Level		
No Reading Abilities	-0.13	0.839***
		(0.093)
Indentify Letters	-0.171	0.777***
		(0.289)
Read Words	-0.195	0.653***
		(0.207)
Read Paragraph	0.34	0.013
		(1.110)
Comprehend Story	0.144	0.653***
		(0.111)
Panel B: Parental Education		
0 years	-0.139	0.820***
		(0.095)
1-5 years	-0.05	0.766***
		(0.110)
6-10 years	0.139	0.693***
		(0.128)
11 years and up	0.147	0.319
		(0.271)