WORKING PAPER
NO. 17

Complement or Substitute? The Effect of Technology on Student Achievement in India

June 2008

This report provides an initial assessment of the pilot computer assisted learning program implemented by Gyan Shala. Designed primarily to reinforce students understanding of material presented within the Gyan Shala classes, the project was implemented in an in-school and out-of-school model. Results suggest that the effect of the program critically depends on the method of implementation and thus highlights the importance of considering the relative productivity of learning models when choosing between educational strategies.

WORKING PAPER NO. 17, 2008

WORKING PAPER NO. 17, 2008

Complement or Substitute? The Effect of Technology on Student Achievement in India

June 3, 2008

Prepared by:

Leigh L. Linden[1]
Columbia University, MIT Jameel Poverty Action Lab, IZA

Edited by:

Michael Trucano
infoDev

[1] As is typical of an undertaking of this magnitude, this project would not have been completed without the assistance of a number of people. Of the many who assisted in this research project, a few made especially important contributions. Pankaj Jain and Zalak Kavi of Gyan Shala are responsible for the management of the data collection activities and the intervention. Nandit Bhatt provided very valuable research assistance, both conducting field interviews with teachers and assisting with data analysis. I am indebted to Esther Duflo and Abhijit Banerjee for helpful suggestions and comments. The intervention and evaluation were funded by InfoDev. All errors are, of course, my own, and any correspondence can be sent to me at leigh.linden@columbia.edu.

Abstract: Using a pair of randomized evaluations, I evaluate a computer assisted learning program designed to reinforce students understanding of material presented in class. The program was implemented in both an in-school and out-of-school model allowing me to assess different strategies for integrating the technology into the existing schools. The effect of the program critically depends on the method of implementation. The program was a poor substitute for the teacher delivered curriculum and as a result, the in-school model caused students to learn significantly less than they otherwise would have learned (-0.57 standard deviations). When implemented as a complement to the normal program in the out-of-school model, however, the program generated average gains of 0.28 standard deviations reflecting small positive (but statistically insignificant) gains by most students and large positive gains by the weakest and older students in the class (from 0.4 to 0.69 standard deviations). The results emphasize the importance of understanding how new technologies and teaching methods both interact with existing resources and differentially affect students with different needs and abilities.

Keywords: Education, Development, Computer Assisted Learning

JEL Codes: I21, I28, O15

I. Introduction

Many have considered the use of computers as a prime opportunity to improve the productivity of classrooms. By providing a dynamic learning experience delivered to children on an individual basis, computers could provide an engaging learning experience tailored to the needs of individual students (see for example, Anderson, Boyle, and Reiser, 1985; Schofield, Eurich-Fulcer, and Britt, 1994). Slower students could practice remedial drills and review material that they have yet to master. Stronger students, on the other hand, could move quickly through additional material, improving their understanding unencumbered by the pace of their slower peers. Within this image, computers can deliver educational inputs that teachers alone could not cost-effectively provide.

These promises are particularly important in the developing country context. Although many countries are making substantial gains towards meeting the Millennium Development Goal of universal primary education by 2015, the quality of schools serving most of these populations remains extremely low (Filmer, Hasan, and Pritchett, 2006). In India, for example, where enrollment rates have increased significantly in recent years, a recent countrywide survey of rural children (ASER, 2007) found that only 58.3 percent of children in fifth grade could read at the second grade level. Similarly, only half of 9-10 year old children who are attending school could demonstrate basic numeric problem solving skills.

Unfortunately, the majority of the general evidence on the effectiveness of large-scale efforts to place computing resources in the classrooms is at best ambiguous with most studies finding small if any effects.[2] Angrist and Lavy (2002) assess the impact of an Israeli school computerization program and find no evidence that the program raised students' test scores. They also find some evidence that the program may have hurt fourth grade students' math scores. Goolsbee and Guryan (2002) evaluate a U.S. program designed to subsidize school use of the

[2]Kirkpatrick and Cuban (1998) conduct a survey of the existing education research and conclude that the evidence is at best mixed.

Internet. They find no affect on student performance. Leuven, Lindhal, and Webbink (2004) evaluate the effect of computer use in the Netherlands finding results similar to those of Angrist and Lavy (2002). The exception in this literature is Machin, McNally, and Silva (2006) who assess the effects of ICT investments in England and find a positive impact on student performance, particularly in English.

The major limitation of these studies is that they treat computing resources as a general educational input. So, while these studies may capture the effectiveness of the average application of computing technology, they cannot speak to the potential effectiveness of individual programs. It may be, for example, that there are potentially effective methods of using computers in the classroom that have just not been implemented on a large enough scale to be measured through such general studies.

A few new studies have begun to estimate the effects of individual programs using randomized evaluation techniques, but even here the results are still ambiguous. Dynarski et al. (2007) evaluate several programs in U.S. schools and find no evidence of effects on students' math and reading scores. Rouse and Krueger (2004) evaluate a reading program in several urban schools in a city in the Northeast of the U.S. and find no evidence that the program generates improvements in students' general language or reading skills. Barrow, Markman, and Rouse (2007) evaluate a program designed to teach algebra and pre-algebra in three U.S. urban school districts, finding that the program improved students' math performance on state administered standardized tests by 0.17 standard deviations.

There are also two evaluations that use similar methodology in the context of a developing country. The evidence is more consistent and suggests that the application of technologies that change pedagogical methods can have very large effects. Bannerjee, Cole, Duflo, and Linden (2007) evaluate the effects of a math program in Vadodara, India finding that the program increases student performance by 0.47 standard deviations on average. Similarly, He, Linden, and MacLeod (2008) evaluate an English language program in Maharashtra, India finding that the program improves students' English scores by 0.3 standard deviations on average.

One limitation of these studies is that they do not consider variation in the way that the individual programs interact with existing resources within the classroom. Most of these studies test individual or multiple interventions without considering, for example, whether or not variations of the individual program might better suit the particular schools involved. Because any intervention rearranges the structure of pre-existing inputs within the classroom, these interactions could be particularly significant. For example, if computers substitute for a more productive arrangement of resources, students may learn less than if the same program was used, instead, to complement existing efforts. This variation in relative productivity could explain some of the inconsistency in the existing literature. For example, it might explain why programs that substitute for teacher instruction in developed countries where the quality of instruction is high have provided more ambiguous results while similar programs in less well-functioning school systems, like those in India, have generated stronger and more consistently positive results.[3]

Working with the Gyan Shala program, an NGO in Gujarat, India, I attempt both to evaluate a novel computer assisted learning program and to do so in a way that allows us to assess the effectiveness of different strategies for the implementation of the program. First, unlike these previous studies, I explicitly take the classroom's existing resources into account when implementing the intervention. Operating in a relatively well-functioning network of NGO-run schools, I first test the effectiveness of the program as a substitute to the regular teaching methods by using the computers in a pull-out in-school program and as a complement by using the computers as an out-of-school time program. In addition, unlike most previous studies, I evaluate a program that is explicitly designed to reinforce the material taught in the normal curriculum. In

[3] In fact, both Angrist and Lavy (2002) and Leuven, Oosterbeek, and Webbink (2004) hypothesize that such substitution may be the cause of the few negative estimates they find in their analysis.

other words, unlike previous programs, the Gyan Shala model focuses on changing the way that material is presented to children within the standard curriculum, rather than allowing children to move at their own pace through additional or more advanced material.

Overall, I find that the program as a whole does little to improve students' math scores. However, I find that the method of implementation matters significantly for the effectiveness of the program. When implemented as a substitute (the pull-out version) to the regular inputs, the program proves much less productive, causing students on average to learn 0.57 standard deviations less than they otherwise would. When implemented as a complement to the existing arrangement of inputs (the out-of-school time version), I find evidence that the program is generally effective, increasing students' test scores by 0.28 standard deviations. When implemented as a complement to existing resources, the program also has differential effects on students. Poorly performing students and older students experience gains of 0.4 to 0.69 standard deviations in math scores, significantly more than their higher achieving peers. Finally, I find that, given the costs and average impacts, the pull-out program is as cost-effective as many other interventions evaluated in the developing country context.

These results suggest that researchers evaluating the effectiveness of new teaching techniques and classroom resources need to consider carefully the way that the new inputs will interact with existing resources and whether that change meets the specific needs of individual types of students. For education and development policy, the results emphasize the importance of conducting rigorous evaluations of educational programs since even small changes to the learning environment can cause significant declines in student performance.

The remainder of the paper is organized as follows. Section Two provides a brief overview of the intervention and the methods of implementation. Section Three provides a description of the general research design and statistical methodology, and Section Four provides the statistical results of the evaluation, and in Section Five, I assess the cost effectiveness of the out-of-school time intervention relative to other education programs that have been assessed in developing countries. I conclude in Section Six.

II. Gyan Shala and the Gyan Shala CAL Program

A. Gyan Shala

Gyan Shala is a project of Education Support Organization, an Indian NGO that aims to develop a new system of solutions for delivering basic-primary education to first through third grade children in poor rural and urban families. The organization attempts to target families whose children would otherwise attend poorly performing public schools or no school at all. The program is based in the state of Gujarat in Western India.

The core innovation of the Gyan Shala learning model is the use of a rigorous, tightly scripted curriculum and the employment of women from the local communities. Like many other programs that employ similarly skilled men and women for educational purposes (see for example the Balsakhi program in Bannerjee, Cole, Duflo, and Linden, 2007), these women would not meet the requirements for employment in the government public schools. The minimum requirement is that teachers pass at least the 10^{th} grade.

Recruited teachers are trained prior to the beginning of the academic year. Gyan Shala trains these women in a very carefully constructed curriculum that proscribes the teachers' activities in 15 minute blocks of time. The material covers all of the basic requirements of the official state curriculum (the Minimum Levels of Learning). The school runs the duration of the normal Indian academic year, but students attend school for only three hours a day. Combined with a careful oversight system and Gyan Shala supplied books and learning materials, this system offers a high-quality, well-structured learning environment at a cost of $40 per student per year.

From the perspective of the student, Gyan Shala emphasizes regular participation and academic achievement. Gyan Shala is run as a non-profit organization, and the fee structure is meant to encourage families to take the children's education seriously. Despite actual costs of $40 per student per year, students are charged only 30 Rupees per month (about $0.75) and the fee is waived if the family is deemed too poor to pay. Because the schools are run by women from the local community, the teachers are able to interact directly with childrens' parents and more carefully monitor the students' needs. Typically, a child enters grade one at an age of at least five and after completing three years of Gyan Shala classes, the child is expected to join grade four in a public or recognized private school.

Gyan Shala started in 2000, and as of the 2004-05 academic year, was running 165 classes, of which 95 were located in the urban slums of Ahmedabad and the rest in villages in three talukas: Dhragandhra, Patdi, and Halol. Operations in Dhrangadra and Patdi were discontinued after the 2004-05 academic year while the organization's efforts in Ahmedabad and Halol continue.

A formal evaluation of the effects of the Gyan Shala program has not been done. However, the children in Gyan Shala schools seem to thrive in the Gyan Shala learning environment. In March 2003, Gyan Shala students in Ahmedabad were tested by an independent evaluation team using the same testing instrument that was being used to asses public school students in the nearby city of Vadodara. Figures 1 and 2 provide a comparison of the scores of third grade students in Gyan Shala to students in grades 3 and 4 in the Vadodara public schools using the math and language questions respectively. Gyan Shala students significantly outperformed the public school students in every subject except copying, even students who were a full grade ahead This, of course, is insufficient evidence to attribute students' performances to Gyan Shala's curricular innovations. Without knowledge of students' counter-factual participation in other education programs, it is impossible to determine whether or not the children would have done equally well in other learning environments. The evidence does, however, demonstrate that the students who attend Gyan Shala's schools seem to be learning significantly.

B. Gyan Shala CAL Program

The Gyan Shala Computer Assisted Learning (CAL) program is designed to complement the students' in-class experience while allowing the student to work independently. The goal of the CAL project is to ensure around one hour of daily computer practice to each child at an annual cost of five dollars per child, exclusive of the cost of power. Two factors help Gyan Shala achieve this cost target. First, Gyan Shala obtained donations of old, used desktops (most with Pentium I processors and a few with Pentium II and III processors) from various sources. Second, the software is designed to facilitate two users at a time by splitting the screen in half vertically and displaying two sets of worksheets simultaneously. Because one child uses the keyboard and one the mouse, each child can work independently of the other.

The computers are organized at a table in the common classroom in groups of four. Each child is allocated to a particular computer, ensuring that no conflicts arise among children about who will work on which machine. Since the computers are used in the slums and interior villages, the project must cope with chronic disruption in electricity supply. To accommodate these power outages, Gyan Shala also supplied each classroom with battery-operated uninterrupted power supplies at each location; this power supply is capable of sustaining the bank of four computers for about seven hours.

Unlike self-paced programs, the Gyan Shala CAL program is designed to complement the days' math curriculum. The software is designed so that children require no support from the teacher. The role of the teacher is confined to switching on the power supply and the computers and to allow different batches of eight children to work on the computers during their allocated time. The schedule of exercises to be presented to the children is drawn up to match the particular day's specified learning schedule, although this matching is only approximate. Two parts of the screen, typically, would present two different exercises (of twenty to thirty worksheets) to limit children's ability to copy from one another. The program supports most of the second and third grade math curriculum.

C. Implementation

The study evaluates the implementation of the CAL program in two different years. The first year of the study assessed the implementation of the CAL program in 23 schools located in two localities, Patdi and Dhrangadra, during the 2004-05 academic year. The second year of the study tracked the implementation of the program in 37 schools located in Ahmedabad and Halol which, relative to Patdi and Dhrangadra, are more urban environments.

While the basic program was used in all areas, the program was implemented as both an in-school and out-of-school program.[4] In the first year when project was implemented in Patdi and Dhrangadhra, the program was run in the latter half of the academic year as an in-school program. Students attended the Gyan Shala schools for the normal three hour period, but worked on the computer-based worksheets instead of participating in the structured Gyan Shala classroom curriculum.

During the second year when the program was implemented in Ahmedabad and Halol, the program was implemented as a year-long out-of-school program. Each location typically ran two classes in a day, one after the other. The students would arrive either before or after school depending on the shift of their class. When one class was going through its normal three hour daily schedule, the children from other class took turns working on the CAL package. In this way, the program supplemented rather than replaced the core Gyan Shala curriculum.

III. Research Design

The primary challenge of assessing the causal impact of changes in teaching method on students' learning outcomes is the potential selection of schools and students into treatment by unobservable characteristics correlated with academic performance. To eliminate this problem, I implemented a random assignment evaluation design in which individual Gyan Shala schools

[4] Gyan Shala planned to implement the same out-of-school program in both years of the study. However, the local administrators of the program in Patdi and Dhrangadhra decided to implement the program on an in-school basis instead.

were randomly assigned either to receive the CAL program or to use only the standard Gyan Shala curriculum. The random assignment of the treatment to schools eliminates the possibility that treatment assignment is correlated with either observable or unobservable school and student characteristics.

A. Sample

Table 1 provides a simple tabulation of the schools and students in our sample. The sample consists of the students attending 60 schools[5] in four locations in Gujarat: Ahmedabad, Halol, Dhrangadhra, and Patdi. For each year of the study, students within the schools were identified based on the cohort of students who took the Gyan Shala final exam at the end of the prior academic year. Any student eligible to enter grades two or three during the year of the study were included within the sample. This included all students in grade two and all passing students in grade one. To minimize the potential differences between treatment and control schools, I stratified the random assignment of schools by the average normalized test score within each year of the study.

During the 2004-05 academic year, the 23 schools in Patdi and Dhrangadhra were randomly assigned to either a treatment group that received the in-school version of the intervention for grades two and three or a control group that did not. Students were identified based on the results of their end-of-year exams in April 2004. The stratified randomization resulted in a treatment group of 11 schools with 392 students and a control group of 12 schools with 387 students.

In the second year of the study, 37 additional schools from Ahmedabad and Halol were added to the study. Students were identified based on their end-of-year exams in April 2005 of the previous academic year. The schools were randomly assigned either to a group that would receive the out-of-school intervention for students in grades two and three during the 2005-06 academic year or a group that would experience only the standard Gyan Shala curriculum. As outlined in Table 1, 19 schools containing 682 students were assigned to the treatment group and 18 schools with 695 students were assigned to the control group.

[5] Sixty-two schools were originally considered for inclusion in the sample. However, two of these schools (one in the first year and one in the second year) were clear outliers in the distribution of average test scores by school. Students in both schools scored on average over a half a standard deviation lower than the school with the most similar score. These schools were randomly assigned separately from the rest of the sample, and their inclusion in the sample does not change the results presented in Tables 3-9.

Overall, the two randomizations resulted in a balanced distribution of schools and students between the research groups for both years. In the combined sample, thirty schools were assigned to the treatment group and thirty were assigned to the control group. This included 1,027 students in the treatment group and 1,082 students in the control group.

B. Data

Three types of data were available for analysis: students' math and language scores in April of the academic year prior to the study (baseline test), students' math and language scores in April of the academic year of the study (follow-up test), and basic demographic information. Baseline assessments of students' performance and students' demographic characteristics allow for a comparison of the research groups to assess their similarity and to gauge the degree to which I can attribute any subsequent differences in student performance to the provision of the treatment. Student knowledge of math as measured in the follow-up test conducted in April of the year of the study then allow for a direct comparison of the treatment and control groups.

Since the CAL program was designed to enhance the students' understanding of the math portion of the Gyan Shala curriculum (which closely follows the official state curriculum), student performance was measured using the exams administered by Gyan Shala to its students. These tests were developed to assess students' knowledge of the major math and language subjects taught by Gyan Shala teachers during the year. Separate exams were administered in grades two and three to allow for differences in difficulty and variation in the actual subjects covered. The format of the baseline and follow-up exams also differed though the format for each was the same in each year of the study. To facilitate comparison between the various versions of the exams, I normalize the scores on each exam relative to the control group distribution for each year of the study and grade.

All data was collected by Gyan Shala staff. To ensure the integrity of the exams, Gyan Shala managerial staff administered the exams directly to children independent of the teachers.

The exams were administered in the individual Gyan Shala classes. The exams were also administered multiple times in each location in an attempt to catch absent students.

Finally, I were also able to identify several demographic characteristics of the children.[6] This included students' gender, religion (Hindu, Muslim, Jain, and Christian), and if Hindu, the students' major caste (Brahmin, Kshatriya, Vaishya, Shudra). Almost 67 percent of the students in the study are Hindu. Twenty percent of the students are Muslim, and 13 percent of the students are either unclassifiable or practice other religions present in India. Of the Hindu students, 34 percent are Kshatriya, 22 percent are Vaishya, and 43 percent are Shudra.

[6] This information was unavailable for all children in Gyan Shala's administrative records. However, almost all Indian names uniquely identify gender, religion, and for Hindu children, caste.

C. Methodology

Because of the random assignment of students to treatment and control groups, the causal effect of the Gyan Shala CAL program can be estimated directly by comparing the performance of students in the treatment group to those in the control group using the follow-up exam. To do this, I employ three statistical models. First, I use a simple difference estimator that compares the average characteristics of the two groups. Second, I use a difference estimator that takes into account variation in students' characteristics within and between the research groups. Finally, I use a difference in differences estimator to compare the attrition rates in the two research groups.

The simple difference estimator has two uses. First, I will use it to compare students using their baseline characteristics to investigate whether any differences exist between the treatment and control groups based on observable student characteristics. Second, I also use this estimator to estimate the raw differences between the two groups on the follow-up test. The estimator takes the following form:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \varepsilon_{ij} \tag{1}$$

Where y_{ij} is the outcome variable for child i in school j, ε_{ij} is a random disturbance term, and $Treat_j$ is a dummy variable for a student attending a class assigned to the treatment group.

Because there are no significant differences between the treatment and control groups in observable characteristics, the effect of the CAL program can be estimated more precisely by taking into account student and school characteristics. This is done using the following specification:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \delta z_{ij} + \varepsilon_{ij} \tag{2}$$

The variable z_{ij} is a vector of student and school characteristics including the baseline measure of students' performance.

Finally, to compare the attrition patterns in each of the research groups, I use a difference in differences estimator that takes the following form:

$$y_{ij} = \beta_0 + \beta_1 Treat_j + \beta_2 Attrit_i + \beta_3 Attrit_i * Treat_j + \varepsilon_{ij} \qquad (3)$$

The variable $Attrit_i$ is an indicator variable for whether or not a child i takes a follow-up test in April of the year of the study. The coefficient on the interaction term, β_3, then estimates the differences in the relative characteristics of attritors versus non-attritors between the treatment and control groups.

Each of these statistical models also has to take into account the fact that students test scores are correlated within schools (Bertrand, Duflo, and Mullainathan, 2003). This correlation is simply due to the fact that students in a school share many characteristics – they are taught in the same way, share classmates, and come from the same community. But the fact that the students' test scores are not independent of each other means that a linear estimator that does not take this into account will overestimate the precision of the treatment effects. The point estimates of the treatment effect will remain unbiased, but the estimate of the variance of the point estimate will be biased downwards. If uncorrected, this downward bias will cause me to reject the null hypothesis too frequently at a specified significance level.

I account for this issue in two ways. First, I follow the majority of the evaluation literature and estimate the OLS model while clustering the standard errors at the unit of randomization (the school) using the Huber-White statistic (Huber, 1967; White 1980, 1982). While this approach has the advantage of being agnostic to the specification of the within-school correlation, it is too conservative because the estimator does not take account of the correlation between observations when estimating the coefficients. To correct for this, I also estimate treatment effects using a nested random effects model with random effects at the school level and, within schools, at the grade level. I estimate this model using Generalized Least Squares. In practice, the difference between these estimators is only relevant for the estimated effects of the out-of-school model in the second year of the study.

For small sample, it is also necessary to correct for the fact that these methods of accounting for within group heterogeneity are only asymptotically consistent. Following Cameron, Gelbach, and Miller (2008), I use the critical values for determining statistical significance from a t distribution with $G-2$ degrees of freedom where G is the number of schools included in the regression (see Table 1). In practice, the number of groups is large enough (especially for the second year results and the results including both years) that these critical values are still very close to the commonly used critical values from the asymptotic distribution.[7]

IV. Results

I organize the results as follows. First, I use the observable baseline characteristics to ensure that the randomization created comparable treatment and control groups. Second, I analyze the attrition patterns and characteristics of students attriting from the baseline to make sure that the initially comparable research groups are still comparable at follow-up. Because of the similarity of the research groups at follow-up, it is then possible to directly estimate the causal effects of the interventions by directly comparing the treatment and control groups using equations one and two.

[7] Specifically, I use the following critical values in two-tailed hypothesis tests. For regressions using all schools (58 degrees of freedom), the critical values are 1.671, 2.000, 2.660 for the ten, five, and one percent significance levels respectively. For regressions including only the first year (21 degrees of freedom), the respective critical values are 1.721, 2.080, and 2.831. For regressions only including the second year (35 degrees of freedom), the respective critical values are 1.690, 2.030, 2.724.

A. Baseline Characteristics

The random assignment of schools either to receive or not to receive the treatment should create two sets of schools with comparable characteristics. To determine whether or not this did indeed occur, I compare the schools and students in the treatment and control group using equation one based on the characteristics of the students available at baseline. Then, I estimate the distribution of the students' math and language scores and compare the entire distribution. All of these estimates suggest that the two groups are, in fact, directly comparable.

The mean differences between the research groups are presented in Table 2. Panel A contains the individual students' test scores. Panel B contains the students' demographic characteristics, and Panel C contains the school-level characteristics. To provide a gauge of the magnitude of the estimated differences, the first column contains a simple correlation between the post-test math scores and the available demographic characteristics using the entire sample of control students. For each of the indicated combination of years, the first and second columns contain the average treatment and control characteristics, and the third column contains the estimated differences.

None of the average differences are large enough to generate significant differences in the students' subsequent follow-up math scores. Considering the combined sample of both years, the differences in the main variable of interest – students' baseline math scores – is less than a hundredth of a standard deviation. The differences in the students' demographic characteristics are also relatively small. The largest difference is the 7.8 and 7.1 percentage point differences in the fraction of students from the Vaishya and Shudra castes respectively, but given the correlation between these characteristics and students' math scores, these differences would generate a difference of only 0.006 and 0.014 standard deviations respectively in the follow-up math test.

Finally, I also compare the characteristics of the students' schools and again, find no average differences. The control schools are slightly larger, but only have 0.267 more students on average than the treatment schools. Similarly, the average math performance of the students in each school is very similar (difference of only 0.013 standard deviations). And finally, the schools appear to have similar diversity of mathematical performance – the standard deviation of the students' test scores differs by less than a hundredth of a standard deviation.

The final two groups of columns in Table 2 then compare the samples of students for the individual years of the study. While the magnitudes of the differences are slightly larger than for the combined samples, none of the differences are statistically significant and the magnitudes are still very small. The largest difference is a difference of 0.071 standard deviations in the second year language test scores. The differences in baseline math scores for each year are both lower than the combined difference of 0.029 standard deviations.

To check for differences between the groups more generally, Figures 3 and 4 contain kernel density estimates of the students' math and language scores respectively. In both cases, the distributions are virtually identical. This and the data from Table 2 suggest that the randomization did indeed create comparable treatment and control groups.

B. Attrition

While the research groups may be initially similar, students who took the baseline exam inevitably fail to take the follow-up exam. Some of these students may have dropped out during the academic year. And some of them may have simply been absent when the testers to administered the follow-up exam. Either way, it is important to compare the students that fail to take the follow-up exam to ensure that the same kinds of students dropped out from each group. If the attrition patterns, on the other hand, were correlated with the administration of the treatment (e.g. there are large differences in the types of students that attrite from each group), then the emergent differences in the treatment and control groups at follow-up would confound the estimate of the causal impact of the treatment.

Table 3 shows the average characteristics of the attriting students in each research group. The table is organized in a similar format to Table 2. Panel A contains the raw attrition rates for each group. Overall, 25 percent of the control students and 23 percent of the treatment students failed to take both parts of the follow-up exam, suggesting that the overall rates of attrition are very similar. Within each year, the difference in attrition rates was slightly larger at 6 and -7 percent in the first and second years respectively. However, even these differences are too small to generate significant differences in the respective samples.

Panel B and C compare the relative attrition patterns in each research group. The average differences in the first two columns contain estimates of the difference between attriting and non-attriting students (attriting students less non-attriting students). The third column then contains the estimated difference between the two estimates, estimated using equation three.

Looking to compositional differences in Panels B and C, it is clear that not only are the same number of children attriting, but the same types of students are also attriting. Poorer performing students are the most likely to drop out of the sample; though on average, the same kinds of students are dropping out of each of the research groups. The largest relative difference in test scores is 0.08 standard deviations for the language test using the entire sample. The differences for the first year are all less than 1.5 hundredths of a standard deviation.

The differences in other student characteristics show similar patterns. For almost all of the characteristics the groups experienced the same attrition patterns. Only two of the estimated differences are statistically significant but again the magnitudes are too small to generate large differences in the resulting sample: a 13 percent difference in the relative proportion of Kshatriya students who drop out in the first year and a 14 percent difference in the relative proportion of second graders who drop out in the second year.

C. One Year Follow-Up, Both Programs

The combined results from the two years of the program suggest that, on average, the program had little effect on students' understanding of math. However, the aggregation masks the significant difference in the performance of the individual forms of the intervention. The in-school intervention seems to have caused students to learn less math than they otherwise would have while the out-of-school program seems to have caused students to learn more. This difference suggests that the computer-based program was a poor substitute for the Gyan Shala teaching environment, but that the program has value as a complement to the basic program.

Panel A of Table 4 contains the comparison of the treatment and control groups using the data from both years of the study. There are three things to note about these comparisons. First, the average performance in the treatment and control groups on the one year follow-up exam are very similar – for both language and math. The first column contains the average value for the control students. The second column contains the average for the treatment group and the third column contains the estimated difference between the two groups. All of the differences are two hundredths of a standard deviation or less.

Second, when the controls are added to the difference equation in column four, the point estimates on the estimated treatment effects barely change. The largest change in the estimated effect is 1.1 hundredths of a standard deviation observed for the difference in language scores. This minimal difference underscores the similarity of the students who took the post-test in the treatment and control groups.

Third, Figure 5 shows the distribution of students' test scores on the math section of the follow-up exam using a kernel density estimator. The results bear out the average differences. The distributions are very similar and the differences that do exist are no larger than those observed in the baseline distributions from Figure 3.

D. In-School Program

Breaking the distributions up by method of implementation yields significant differences in the effects of the two implementations. Panel B of Table 4 contains the results for the first year of the program when the program was implemented on an in-school basis. The results in Panel B document that for the in-school version of the program, students who received the intervention performed worse than students who did not. Overall, students performed 0.48 standard deviations less on the follow-up test overall and 0.57 standard deviations less on the math portion of the test. There is some evidence that the treatment also reduced students' language scores (a possible indirect effect), but the effects are sufficiently small that they are not statistically significant.

Figures 6 and 7 estimate the difference in treatment effects for the entire distribution of students. Figure 6 contains kernel density estimates of the students' respective performances on the math section of the follow-up exam, similar to those of Figure 5. The distribution yields a very clear pattern with fewer students in the treatment group scoring over zero and more students scoring less than zero. Figure 7 contains estimates from a non-parametric regression of students' follow-up tests scores on their baseline scores. This allows for a comparison of post-test scores conditional on students' baseline performance. The results are very consistent showing that across the distribution, treated students seem to do equally worse than their untreated peers.

While large, these estimates are consistent with the positive effects measured in other programs. For example, Banerjee, Cole, Duflo, and Linden (2007) find positive effects of 0.47 standard deviations from a computer-based math program targeted at similar students. Given that this program generated these gains by providing only 2 hours of relatively more productive instruction a week (1 hour outside of school and 1 hour in-school), a decline of 0.57 standard deviations is a plausible effect of substituting a less effective form of instruction for a full third of the school day.

To check the consistency of this result, Table 5 provides the same estimates for the first year of the program on individual math subjects on the second and third grade exams. Because the exams covered different subjects, it is necessary to provide the estimates disaggregated by grade. Results are presented first for grade two and then grade three. The overall estimates for each grade are provided first with the results disaggregated by subject below. It is important to keep in mind that because the estimates have to be made within grade, the sample size in these regressions is much smaller than for the entire sample (329 second graders and 197 third graders).

Despite the smaller sample size, the results generally bear out the overall estimates. The overall average treatment effect for both second and third graders is negative. Younger students seem to fare worse than older students, though the difference is not statistically significant at conventional levels of significance. The individual subjects follow the same pattern with all students generally learning less due to the program, but second grade students learning less than their third grade peers.

Students' performance in each subject declines significantly with the possible exception of single digit multiplication.

Table 6 estimates the differences in math scores while dividing the students by demographic characteristics rather than by subject. For reference, the over all average difference in math scores is provided in the first row. Panel A contains the estimated effects by gender. Panel B divides students by religion. And Panel C divides the students by their performance on the baseline exam. Students are divided into terciles with the weakest performing students in tercile one.

The results are generally the same as the overall averages and are consistent with the differences depicted in Figure 7. Across almost every subgroup of students, the treatment group underperformed the control group. Muslims and students from the Vaishya caste seemed to fare the best, but given the small number of these students, the differences are not statistically significant.

E. Out-of-School Program

Unlike the in-school program, the results from the second year of the program suggest that the added value of the computer assisted learning program in addition to the Gyan Shala curriculum is positive. The average estimates in Panel C of Table 5 show that the effects of the program on math are about 0.27 standard deviations – an estimate that is consistent with those of other programs. However, the statistical significance of this result depends on the assumptions made about the correlation of students' scores within a school and the efficiency of the employed estimator. The difference is not statistically significant under the clustered standard errors, but it is statistically significant at the five percent level using the more efficient random effects estimator.[8]

Figure 8 shows the distribution of students' follow-up test on the math section of the exam. Like Figure 6 the distributions show a sharp departure from Figure 5, but unlike Figure 6, students in the treatment group are more likely to score above zero on the follow-up exam. Figure 9 contains non-parametric estimates of students' follow-up score on their baseline score by research group. Unlike the estimates for the in-school program, these estimates suggest that the program had a more positive effect for students at the bottom of the test score distribution than for students who performed better on the baseline test.

Turning to Table 7, I estimate the effects by subject as in Table 5. These results suggest that the program had a significant effect for students in grade three and little effect on students in grade two. Students in grade three benefited by 0.515 standard deviations overall while students in grade two show an increase of only 0.077 standard deviations. The difference in the effects is statistically significant at the 10 percent level (p-value 0.082). Because this program was self administered, it may be that older students were better able to take advantage of the program. As in Table 5, these results are consistent across the individual subjects.

Table 8 shows the estimated treatment effects disaggregated by demographic characteristics. The organization is the same as Table 6. The results are generally consistent with the non-parametric estimates in Figure 9 – the program has a large statistically significant effects

[8] It is important to note that there are two differences between the random effects estimate and the clustered standard errors. First, the standard error falls from 0.172 in the clustered estimates to 0.166 in the random effects estimates as one would expect given the greater efficiency of the estimator. The random effects estimator, however, is only asymptotically consistent and in what may be a result of small sample bias, the point estimate is 6.7 hundredths of a standard deviation higher under the random effects estimate. However, the significance of the random effects point estimate is not solely driven by the change in magnitude since even the point estimate from the OLS regressions would be statistically significant at the 10 percent level using the standard errors from the random effects estimate.

for poorly performing students. The point estimates show large positive effects for boys (0.398 standard deviations) and Muslim students (0.688 standard deviations). Both of these subgroups of students had a negative average score on the baseline math exam. The treatment effect estimates for students in the bottom tercile of the baseline math distribution confirm this assessment as these students experience a treatment effect of 0.472 standard deviations that is (like those for boys and Muslims) statistically significant at the 5 percent level. This effect is 0.35 standard deviations larger than that of the strongest students (p-value 0.065) and is 0.29 standard deviations larger than the effect for terciles two and three combined (p-value 0.054).

To check the robustness of this result, I re-estimate the effects of the program on individual subjects (as in Tables 5 and 7) using only students in the bottom tercile. The results are presented in Table 9. Unlike the results for all students (Table 7), students in both grades show a similarly strong response to the treatment with students in the second grade increasing their score by 0.498 standard deviations and students in the third grades increasing their score by 0.524 standard deviations. The point estimates for all but one of the subjects is positive (number ordering is small and negative), and the gains seem to be reasonably well distributed across subjects.

Overall, these results suggest that the out-of-school model of the program significantly improved the math performance of older students and the poorest performing students in the Gyan Shala schools on almost all subjects taught. The effect for older students may results from the self-administered nature of the program while the heterogeneity in the treatment effect by ability may reflect the overall structure of the program. Because the learning model was designed to reinforce the lessons taught by the teacher rather than to allow students to move forward at their own pace, it seems reasonable that a student who already understood the material based on the classroom lectures would gain little from practicing on the computers outside of class. However, students who did not completely comprehend the material apparently found this additional instruction to be helpful. This is consistent, for example, with the results of He, Linden, and MacLeod (2008) who find that stronger performing children benefit relatively more from a self-paced English language program while weaker students benefit more from structured games provided by the teachers that reinforce existing lessons.

V. Cost-Effectiveness

By considering the cost of the overall average change in student test scores, I can compare the cost-effectiveness of the out-of-school program to other programs that have been evaluated. At an average effect of 0.27 standard deviations, the projected cost effectiveness of the intervention is $1.85 dollars per tenth of a standard deviation. However, because all of the original hardware was donated by companies, this only includes the cost of repairing the donated computers. To consider the true cost-effectiveness of the project, I must consider the actual value of the donated computers. Unfortunately, the age of the computers makes it difficult to estimate both the cost and expected life of the machines. Since each computer is used by an estimated nine children, the cost of the program will increase by $3.70 per child or $1.37 per child-tenth of a standard deviation for every $100 spent on a computer assuming that the computers depreciate over three years. So, at $100-$200 per computer (which given the age of the computers is reasonable), the cost per tenth of a standard deviation would increase to $3.22 to $4.59 per tenth of a standard deviation.

Based on these estimates, the program is on par with other interventions considered for improving student performance in developing countries. It is more cost effective than the $7.60 per tenth of a standard deviation math-based computer assisted learning program evaluated by Bannerjee, Cole, Duflo, and Linden (2007), and it is as cost effective as a girls scholarship program ($1.77 to $3.53 per child per tenth standard deviation, Kremer Miguel Thornton 2007), cash incentives for teachers ($3.41 per student per tenth standard deviation, Glewwe et al. 2003), and textbooks ($4.01, Glewwe et al. 1997). It is, however, less cost-effective than a remedial education program ($1 per tenth of a standard deviation, Banerjee, Cole, Duflo, and Linden,

2007) and an English teacher training program ($0.24 per tenth of a standard deviation, Linden, He, MacLeod, 2008).

VI. Conclusion

The effect of the Gyan Shala Computer Assisted Learning Program on students' math scores depends on the method in which the program is implemented. Compared to the apparently productive learning experience students encounter in the normal Gyan Shala curriculum, the Computer Assisted Learning program is a poor substitute. As a result, students who experience the program instead of the normal curriculum perform worse than students who do not receive the treatment. In this model of the program, students receiving the program performed on average 0.57 standard deviations worse in math than students who did not receive the program.

The out-of-school model, however, performs differently. The program has an average effect on all children of 0.28 standard deviations in math. This average reflects small positive (but statistically insignificant) gains for most students and large positive gains of 0.47 to 0.68 standard deviations for the poorest performing students and older students. The difference in the magnitude of the treatment effect for stronger and weaker students seems to reflect the design of the program which emphasized reinforcement of material that students had already learned rather than self-paced discovery of subjects not yet covered in the regular Gyan Shala classes.

More generally, these results emphasize the importance of considering both relative productivity of learning environments when choosing educational interventions and the effects those differences will have on different types of students. Decision-makers must consider not just whether a program works, but rather how well the program works relative to what students would otherwise experience. And they must consider whether those differences are appropriate for individual learners. As this evaluation demonstrates, the format of the program can make the difference between providing needed assistance to weak students and generally causing all students to learn less than they would have learned without the intervention.

VI. Bibliography

Angrist, Joshua and Victor Lavy. (2002). "New Evidence on Classroom Computers and Pupil Learning," *The Economic Journal*. 112(482): 735-765.

Anderson, John, C. Franklin Boyle, and Brian J. Reiser. (1985). "Intelligent Tutoring Systems," *Science*. 228(4698): 456-462.

Annual Status of Education Report (ASER). (2007). http://www.pratham.org/aser07/ASER07.pdf. Accessed: May 19, 2008.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. (2007). "Remedying Education: Evidence from Two Randomized Experiments in India," *Quarterly Journal of Economics.* 122(3): 1235-1264.

Barrow, Lisa, Lisa Markman, and Cecilia Rouse. (2007). "Technology's Edge: The Educational Benefits of Computer-Aided Instruction," *Federal Reserve Bank of Chicago Working Paper 2007-17.*

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. (2004) "How Much Should We Trust Difference in Differences Estimates?" Quarterly Journal of Economics 119(1): 249-275.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2007). "Bootstrap-Based Improvements for Inference with Clustered Errors," Forthcoming *The Review of Economics and Statistics*.

Dynarski, Mark. (2007). "Effectiveness of Reading and Mathematics Software Products: Findings from the 1st Student Cohort." *Mathematica Report.* http://www.mathematica-mpr.com/publications/redirect_pubsdb.asp?strSite=PDFs /effectread.pdf. Accessed: May 19, 2008.

Filmer, Deon, Amer Hasan, and Lant Pritchett. (2006). "A Millennium Learning Goal: Measuring Real Progress in Education." *Center for Global Development Working Paper Number 97.*

Glewwe, P., Kremer M., Moulin, S. (1997). "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya," unpublished working paper.

Glewwe, P. Nauman, I., & Kremer M. (2003). "Teacher Incentives," *National Bureau of Economic Research Working Paper # 9671*.

Goolsbee, Austan and Jonathan Guryan. (2006). “The Impact of Internet Subsidies on Public Schools,” *Review of Economics and Statistics*. 88(2): 336-347.

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

Kirkpatrick, Heather and Larry Cuban. (1998). “Computers Make Smarter Kids – Right?” *Technos Quarterly*. 7(2): 1-11.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. (Jan. 2007). "Incentives to Learn." *Harvard University Working Paper*.

Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek, and Dinand Webbink. (2007). “The Effect of Extra Funding for Disadvantaged Pupils on Achievement,” *Review of Economics and Statistics*. 89(4): 721-736.

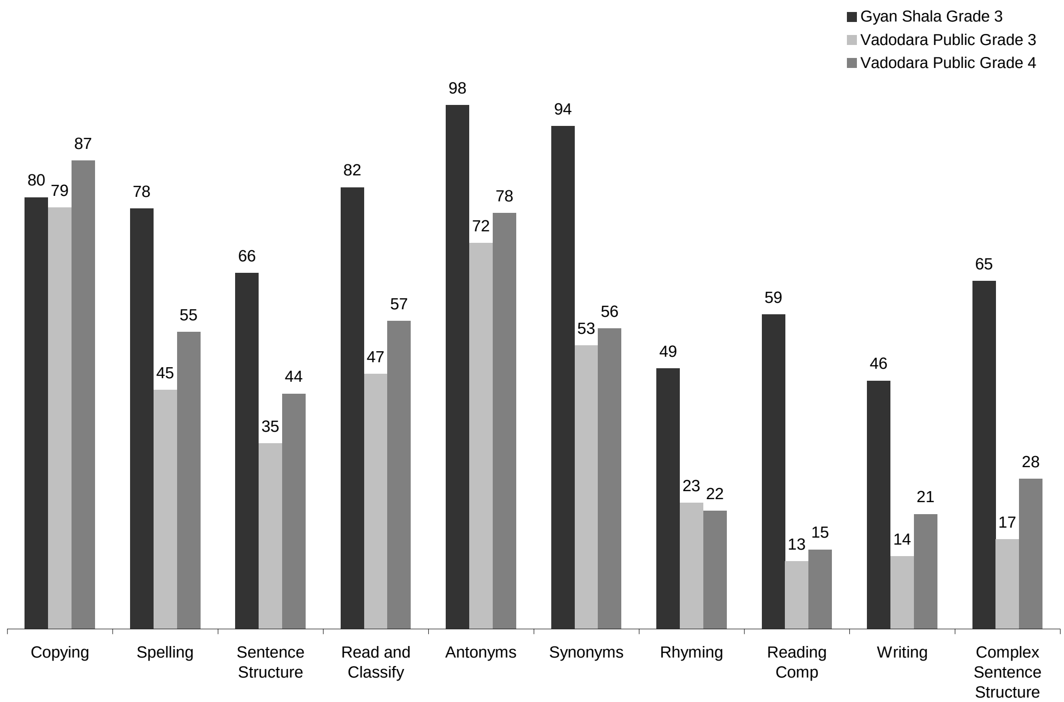
Machin, Stephen, Sandra McNally, and Olmo Silva. (2006). “New Technology in Schools: Is There a Payoff?” *The Economic Journal*. 117(522): 115-167.

Rouse, Cecilia and Alan Krueger. (2004). “Putting Computerized Instruction to the Test: A Randomized Evaluation of a “Scientifically Based” Reading Program,” *Economics of Education Review*. 23(4): 323-338.

Scholfield, Janet, Rebecca Eurich-Fulcer, and Cheri Britt. (1994). “Teachers, Computer Tutors, and Teaching: The Artificially Intelligent Tutor as an Agent for Classroom Change,” *American Educational Research Journal*. 31(3): 579-607.

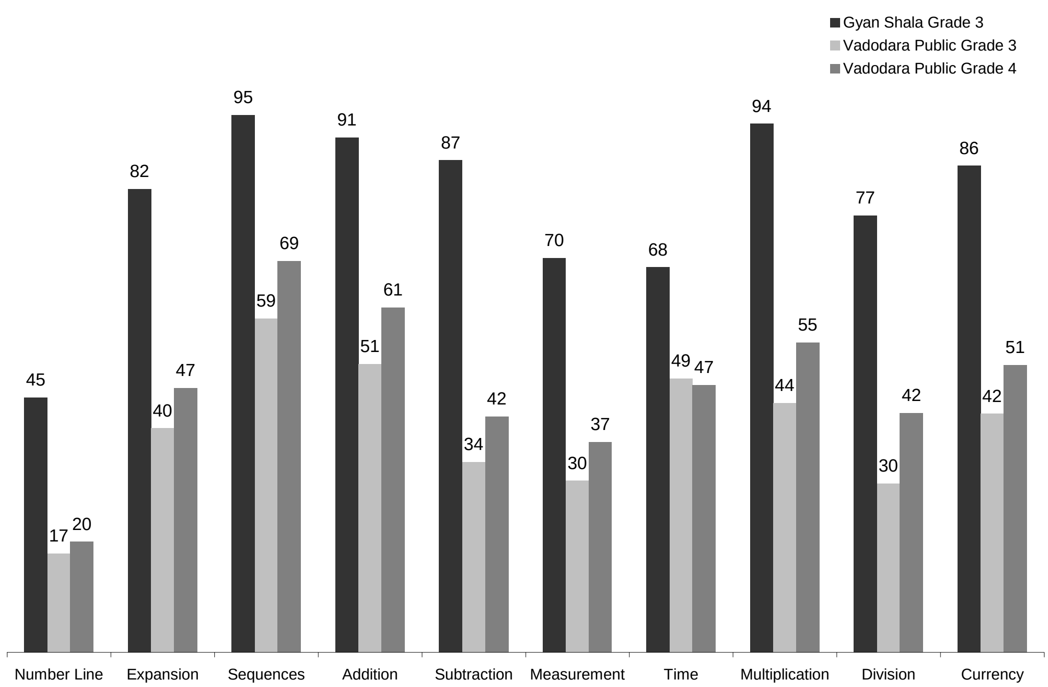
Wenglinsky, H. (1998). “Does it compute? The Relationship Between Educational Technology and Student Achievement in Mathematics.” Princeton, NJ: Educational Testing Service.

Figure 1: Gyan Shala and Public School Student Performance, Language



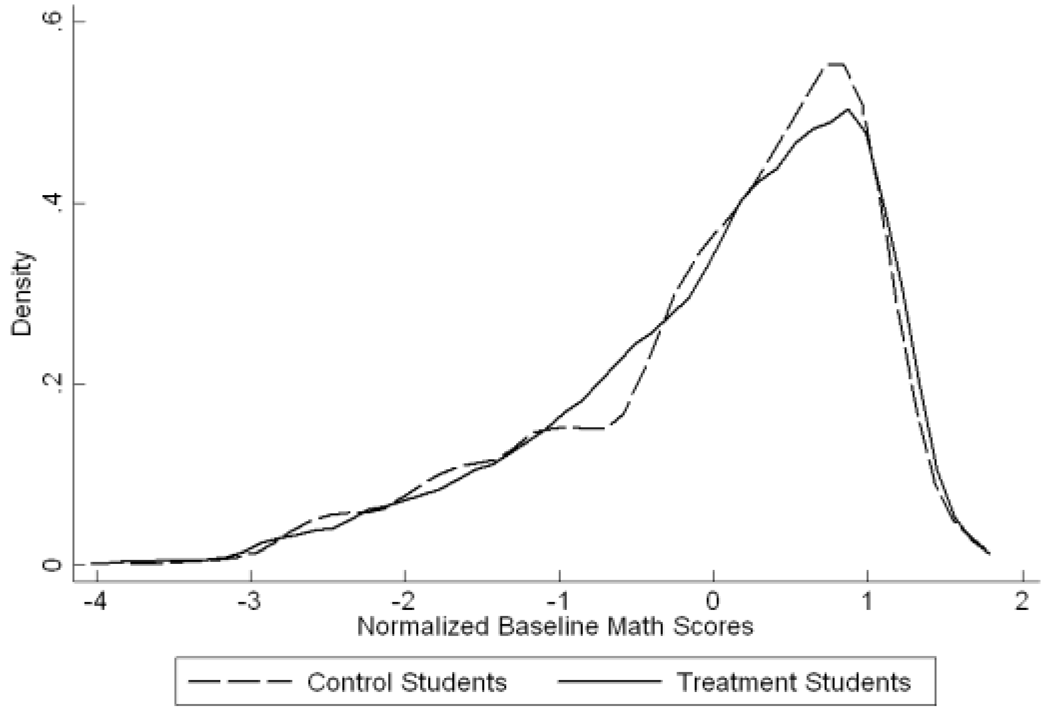
Note: Comparison of Gyan Shala third graders and students from the third and forth grade of the Vadodara Municipal Corporation (MNC) public school system. All scores are percentage of correct answers on the indicated subject.

Figure 2: Gyan Shala and Public School Student Performance, Math



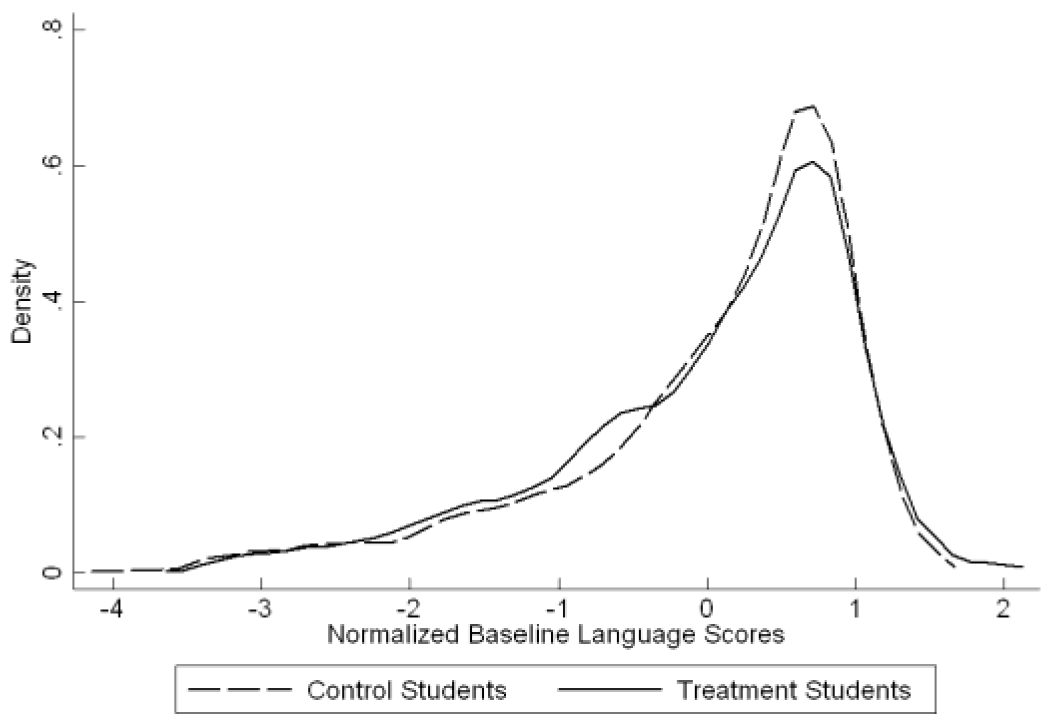
Note: Comparison of Gyan Shala third graders and students from the third and forth grade of the Vadodara Municipal Corporation (MNC) public school system. All scores are percentage of correct answers on the indicated subject.

Figure 3: Distribution of Normalized Math Scores at Baseline



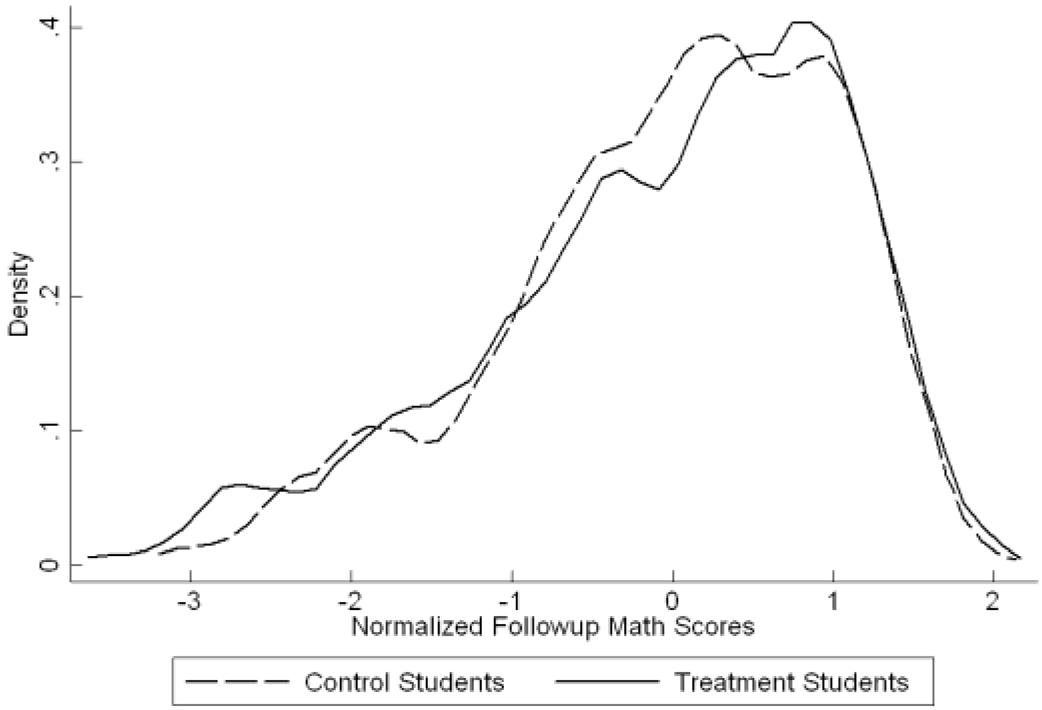
Note: Kernal density estimate of the distribution of baseline math scores for the indicated research group using observations from both years of the study. Bandwidth set to 0.3 standard deviations.

Figure 4: Distribution of Normalized Language Scores at Baseline



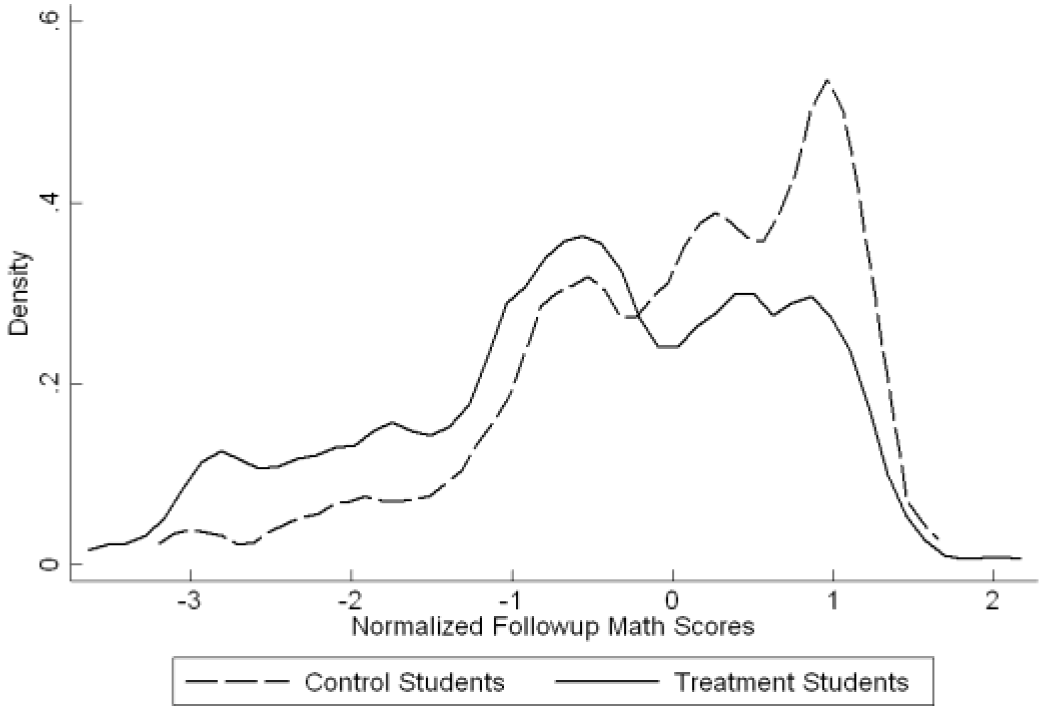
Note: Kernal density estimate of the distribution of baseline language scores for the indicated research group using observations from both years of the study. Bandwidth set to 0.3 standard deviations.

Figure 5: Distribution of Normalized Math Scores at Follow-Up



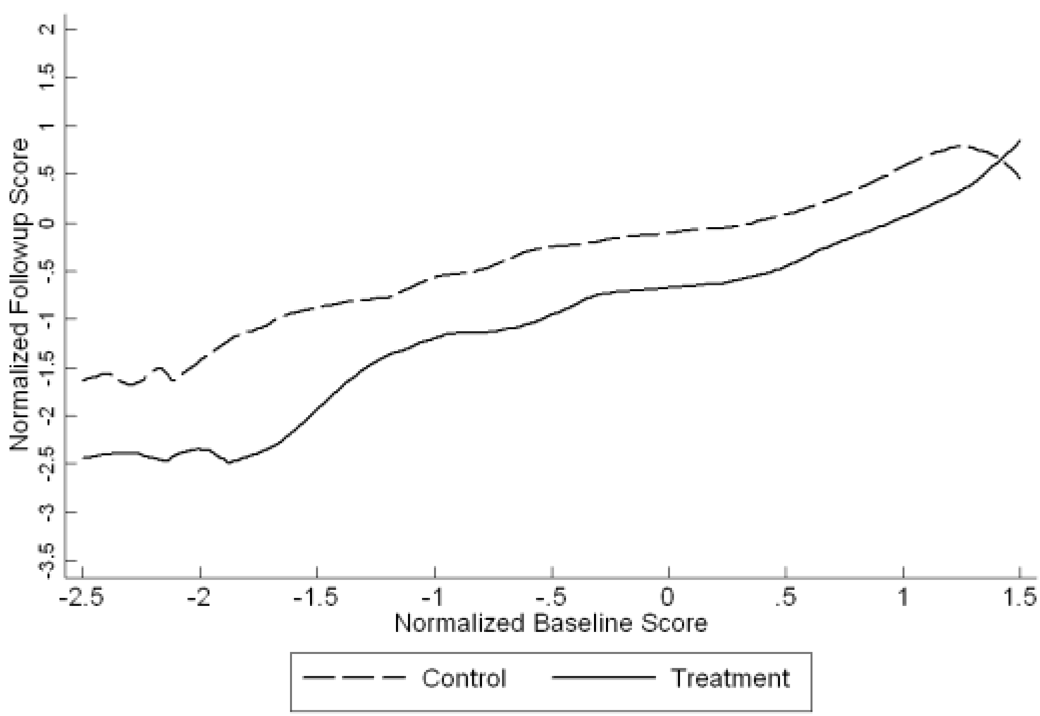
Note: Kernal density estimate of the distribution of follow-up math scores for the indicated research group using observations from both years of the study. Bandwidth set to 0.3 standard deviations.

Figure 6: Distribution of Normalized Math Scores at Follow-Up, In-School



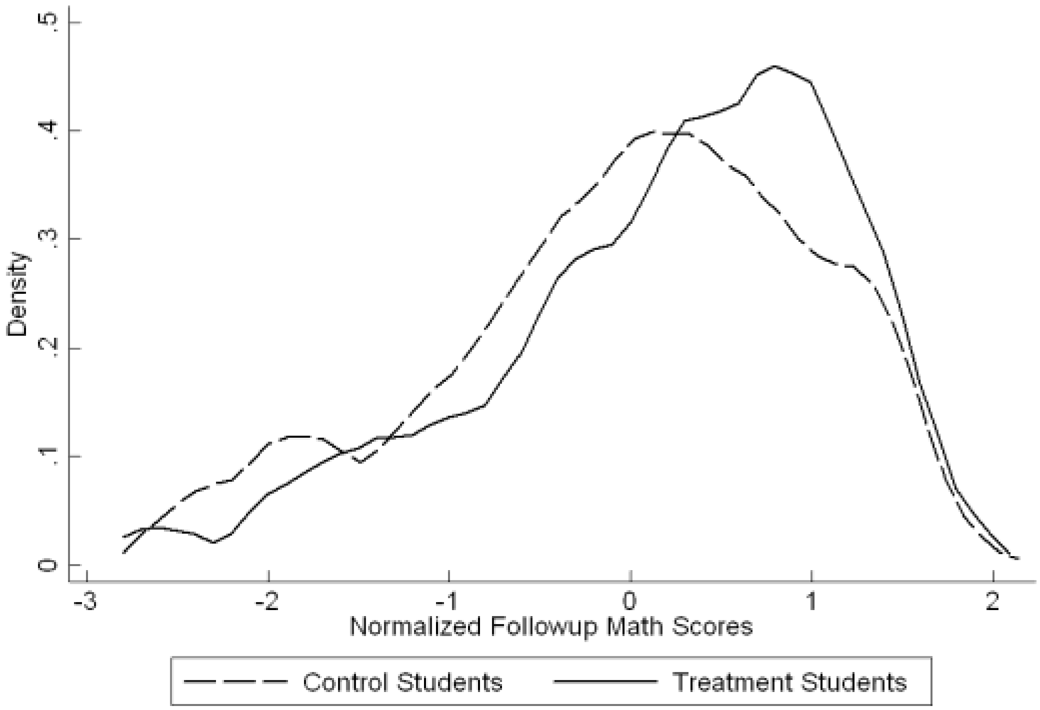
Note: Kernal density estimate of the distribution of follow-up math scores for the indicated research group using observations from the first year of the study. Bandwidth set to 0.3 standard deviations.

Figure 7: Follow-Up Math Scores by Baseline Math Scores, In-School



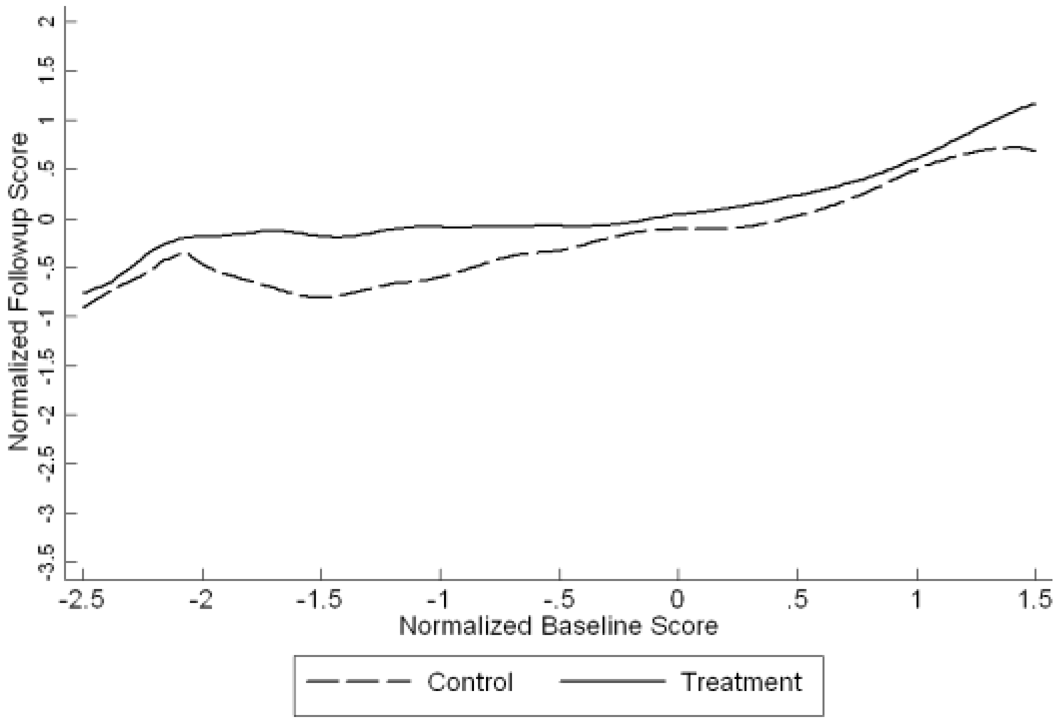
Note: Local linear polynomial estimates of the relationship between the normalized math follow-up scores and normalized baseline scores for the first year of the study. Bandwidth set to 0.5 standard deviations.

Figure 8: Distribution of Normalized Math Scores at Follow-Up, Out-of-School



Note: Kernal density estimate of the distribution of follow-up math scores for the indicated research group using observations from the first year of the study. Bandwidth set to 0.3 standard deviations.

Figure 9: Follow-Up Math Scores by Baseline Math Scores, Out-of-School



Note: Local linear polynomial estimates of the relationship between the normalized math follow-up scores and normalized baseline scores for the second year of the study. Bandwidth set to 0.5 standard deviations.