Improving the Quality of Education in India: Evidence from Three Randomized Experiments

Abhijit Banerjee, Shawn Cole, Esther Duflo and Leigh Linden*

March 10, 2003

Abstract

This paper presents the results of a two-year randomized evaluation of a large scale remedial education program, conducted in Mumbai and Vadodara, India, and the preliminary results of a randomized evaluation of a computer assisted learning program in Vadodara. The remedial education program hires young women from the community to teach basic literacy and numeracy to children who reach standard three or four without having mastered these competencies. The program, implemented by a NGO in collaboration with the government, is extremely cheap (it cost 5 dollars per child per year), and is easily replicable: It is now implemented in 20 Indian cities, and reaches tens of thousands of children. We find the program to be very effective: On average, it increased learning by 0.15 standard deviations in the first year, and 0.39 in the second year. The gains are the largest for children at the bottom of the distribution: Children in the bottom third gain 0.18 standard deviations in the first year, and 0.59 in the second year. The results are very similar in the two standards, and in the two cities. At the margin, extending this program would be 4.5 to 6 times more cost effective than hiring new teachers. The preliminary results of the computer assisted learning program, which is planned to be widely implemented in India, are less impressive:

*PRELIMINARY AND INCOMPLETE. This project is a collaborative exercise involving many people. Foremost, we are deeply indebted to the Pratham team, who made the evaluation possible and put up with endless changes in plans and new requests: Rukmini Banerjee, Lekha Bhat, Madhav Chavan, Shekhar, Rajashree Kabare, and Pratima. We thank Jim Berry and Marc Shotland, for their excellent work coordinating the fieldwork in Vadodara. They were helped by Nandit Bhat and many others. For financial support, we thank the ICICI corporation, the World Bank, and the MacArthur Foundation Network on the Costs of Inequality.

On average, the program increases test scores by an insignificant 0.10 standard deviations. The effect is higher (and significant) in schools where the remedial education program is also present. On the basis of these estimates, extending the computer assisted learning program would appear less cost effective than hiring new teachers.

1 Introduction

There has been a lot of interest recently in the question of how to effectively deliver education to the poor in developing countries and a corresponding apparition of high quality research on the subject. A lot of the research focuses on the effects of reducing the cost of schooling, with the view that the important goal is to get the children into school. Examples of this kind of work include Banerjee, Jacob and Kremer (2002) on school meals in India, Duflo (2001) on school construction in Indonesia, Glewwe, Kremer and Moulin (1997a) on school uniforms in Kenya, Spohr (1999) on compulsory schooling laws in Taiwan and Vermeersch (2002) on school meals for pre-schoolers in Kenya. The primary metric by which success is judged in these studies is attendance, and in each of these cases there was found a significant impact.

Are students also learning measurably more as a result of these interventions? There is no obvious reason why they would. The influx of new students probably makes learning harder for the children who were already in school, simply because there are more demands on existing resources.[1] And while the newcomers will presumably learn more, just by the fact that they are now attending school, it is not clear that there is anyone with whom we could compare them.

At the other extreme are interventions that focus directly on improving test scores for students who are already in school. These are interventions where students are explicitly rewarded for doing well on tests: Angrist, et al, (2002) study a program in Colombia that offers private school vouchers to students who keep their scores above a certain level. A recent study by Kremer, Miguel and Thornton (2002) looks at the impact of offering scholarships to students in Kenya who do well on a standardized test. Both studies find an impact on test scores, though in such cases the existence of an impact is perhaps less interesting than whether the impact is commensurate with the money spent.

Perhaps the most interesting case is the one in between: Interventions that purport to improve the quality of the learning experience, but for which no evidence exists that they actually do improve learning. Examples include increasing the teacher-student ratio (Banerjee, Jacob and Kremer, 2002), subsidized textbooks (Glewwe, Kremer and Moulin, 1997), free flip-charts (Glewwe, Kremer, Moulin and Zitzewitz, 1997); and then the interventions that improve the

[1] Indeed this is what Banerjee, Jacob and Kremer (2002) find for mid-day meals, and Glewwe, Kremer and Moulin (1997) find for a program that offered both free textbooks and free school uniforms.

health of school children (for example, deworming, as in Kremer and Miguel, 2002), incentives for teachers (Glewwe, Kremer and Moulin, 2002), and blackboards and other school inputs (Chin, 2001), etc. If these programs affect attendance, it is only because school quality goes up. In such cases, we ought to expect an improvement in test scores among those who were already in school.

Nevertheless, it is notable that relatively few of the studies from developing countries report a positive impact on test scores for those who were already in school.[2] Moreover, the idea that there really was no impact on their education cannot be ruled out *a priori*: The quality of teaching in these schools often leaves much to be desired. Or it could even be the case that the children do not learn because they do not want to: The returns are just not high enough.

This paper reports on a randomized evaluation of two interventions from India, both intended to improve school quality and test scores, and both implemented under the aegis of Pratham, a Bombay-based Non-Governmental Organization. Both were implemented in urban slums, where most children of the relevant age do attend school, however desultorily, so that the main impact is on the learning environment of existing students. The first program is a supplemental teaching program, where a fraction of the third or fourth standard (grade) is pulled out of regular instruction and given remedial teaching with the goal of helping them catch up with the rest. The intervention is motivated by the well-known fact that children often drop out because they fall behind and then feel lost in class.

The second intervention seeks to help children in the fourth standard improve their mathematics skills by playing specially designed educational games on the computer. The goal was to replace the rather passive learning environment in most Indian schools with a more active one—to make learning less boring.

[2]The one exception we of we are aware is the study of a program that provides incentives for teachers in Kenya that is reported in Glewwe, Kremer and Moulin (2002b), though even in this case the authors seem to be somewhat disappointed by the lack of a more robust impact. Chin (2001) finds that Operation Blackboard in India did increase school completion rates for girls, which implies that there must have been an increase in test scores, but she cannot tell whether those who would have completed school in any case learn more as a result of the intervention. Vermeersch (2002) also finds an impact on test scores of a school meals program in schools where the teachers were trained, but she too cannot distinguish between those who were already in school and the newcomers.

The evaluation of the supplemental-teaching (balsakhi) program offered an opportunity to implement an evaluation design that is often recommended but rarely, if ever, utilised.[3] First, it was a randomized evaluation. We can therefore be relatively confident of the absence of confounding factors. Second, it was an evaluation of a program that had already clearly demonstrated the ability to scale up, as the description below will make clear. In other words, there is no danger that what we are evaluating cannot be reproduced elsewhere. Third, we simultaneously carried out randomized evaluations of the program in two different cities, where each project was run by a different set of people. This reinforces our confidence in the external validity of these results. Finally, we have results that cover two years and several tests, making it less likely that the results reflect the fact that the program is new or the effect of initiating an evaluation.

We find that the balsakhi program did have a substantial impact, and that impact is remarkably stable across years and cities, especially when we take into account the instability of the environment—there was a major riot and a catastrophic earthquake while the program was running. Moreover, the weaker students, who are the target of the program gain the most. This makes it clear that it is possible to improve test scores substantially at very low cost.

The computer assisted learning (CAL) program, on the other hand, is brand new, although there have been some pilot projects along these lines previously. It is much more expensive than the Balsakhi program, but nevertheless the idea is being widely discussed in India and elsewhere. In this sense, an evaluation is clearly necessary. However, it is not clear that the *modus operandi* for the CAL program has been fully worked out, and the results may reflect deficiencies of the design rather than intrinsic limitations of the concept.

With this important caveat, the results from the randomized evaluation are disappointing. The program seems to have only a modest (and insignificant) impact on test scores, though the children seem to be very happy to play the games on the computer.

The rest of this paper describes the programs and our evaluations in more detail before presenting the results.

[3] We are unaware of any other evaluation of an educational program that meets these criteria.

2 The Programs

2.1 Remedial Education: The Balsakhi Program

Pratham was established in Mumbai in 1994, with support from UNICEF, and has since then expanded to several other cities in India. Pratham now reaches over 121,000 children in 20 cities in India, and employs about 10,000 individuals. Pratham works closely with the government: Most of its programs are conducted in the municipal schools, and Pratham also provides technical assistance to the government.

One of Pratham's core programs is a remedial education program, called the balsakhi program. This program, in place in the municipal schools, provides a teacher (usually a young woman, recruited from the local community, who has herself finished secondary school) for children identified as falling behind their peers. While the exact details vary depending on local conditions, the typical instructor meets with a group of approximately 15-20 children in the morning for two hours, and with another group of the same size in the afternoon. Instruction focuses on the core competencies the children should have learned in the second and third standards, primarily basic numeracy and literacy. The instructors are provided with a standardized curriculum that was developed by Pratham. They receive two weeks of training at the beginning of the year and ongoing reinforcement while school is in session. The balsakhi program is in place in most of Pratham's sites. It was started in Mumbai in 1994, and then expanded to Vadodara in 1999.

According to Pratham, the main benefit of the program is to provide individualized, non-threatening, attention, to children who are lagging behind in the classroom, and are not capable of following the standard curiculum. Children may feel more comfortable with women from their own communities. As the balsakhi's class size is relatively small, she may tailor this curriculum to their specific needs. Furthermore, because Pratham's program takes children out of the classroom, it may even benefit children who were not directly targeted by the intervention. Removing children from the classroom for two hours means the effective student-teacher ratio in the main classroom drops. Additionally, if the balsakhis are indeed effective, even when the children are returned to the main classroom, the teacher may not need to keep re-teaching remedial material.

An important characteristic of this program is the ease which with it can be scaled up. Because Pratham relies on local personnel, trained for a short period of time, the program is very low-cost (each teacher is paid 500 rupees, or 10 dollars, per month) and is easily replicable. There is rapid turnover among the balsakhis (each of them staying for an average of one year, until they get married or get another job), indicating that the success of the program does not depend on a handful of very determined and enthusiastic individuals. Finally, since the schools provide classroom space, the program has very little overhead or capital costs. The three-year study, nearing completion, was implemented to determine how effective the intervention is.

2.2 Computer Assisted Learning

This program takes advantage of a policy put in place by the government of Gujarat in 2000: The government delivered four computers to each of 100 primary schools in the city (80% of the public schools). A survey conducted by Pratham in June, 2002, suggested that very few of these computers were actually used by children in elementary grade levels. While some schools may have run programs for older students or allowed teachers to use them for administrative tasks, most of the computers remained in their boxes, for want of anyone capable of operating them.

Pratham had previous experience with computer assisted learning, having run a small computer assisted learning program in Mumbai for several years. In particular, they had developed instructional software in the local language, Gujarati. After consultation with the Vadodara Municipal Corporation, they introduced a computer assisted learning program in half of the VMC schools, using the computers already present when possible and replacing or adding computers where necessary.

Pratham hired a team of instructors from the local community and provided them with five days of training. These instructors provided children with two hours of shared computer time per week (two children sharing one computer) – one hour during class time and one hour either immediately before or after school. During that time, the children played a variety of educational computer games chosen because they emphasized some of the basic competencies in the VMC mathematics curriculum.

Pratham designed the program to allow the children to learn as independently as possible. The instructors encouraged each child to play games that challenged the student's level of understanding and, when necessary, helped individual children understand the tasks required of them by the game. All interaction between the students and instructors is driven by the child's use of the various games and at no time do any of the instructors provide general instruction in mathematics.

Schools where the CAL program was not implemented are free to continue to use the computer at their convenience, but our observation is that, except for a small number of schools, they did not start to make use of them for instructional purpose.

3 Evaluation Design

3.1 Sample: Vadodara

In 2000, when Pratham decided to expand their remedial education (balsakhi) program to cover the entire city of Vadorara, they decided to take advantage of the expansion to evaluate the effectiveness of the program in the remaining 98 eligible schools in the city. In November, 2000, they conducted a cognitive test (designed by the Pratham team) of all the children in the third standard. They then hired and trained balsakhis, which were sent to a randomly selected half of the schools in Vadodara. Schools were stratified by medium of instruction, gender, and pupil-teacher ratios. Unfortunately, the school year was disrupted by an earthquake in Gujarat, and children received only a few weeks of instruction between November and March. The first year of the program is best understood as a pilot program.

Starting in July, 2001, the group of schools that had received a balsakhi in the first year of the program received the balsakhi in the fourth standard, and the remaining schools received a balsakhi in the third standard. Children in the standard that did not receive the balsakhi in a given grade form the comparison group for children who did receive the balsakhi.

The program was continued during the school year 2002-2003, with the addition of the 25 remaining primary schools. Schools where the balsakhi was assigned in standard three in the year 2001-2002 were now assigned a balsakhi in standard four, so that children who are now in standard four and are in the treatment group have benefitted from two years of the balsakhi program. Schools where the balsakhi was assigned in standard four in the year 2001-2002 now receive balsakhi assistance for standard three. The new schools were randomly assigned to either group with equal probability in the same way that the original schools were assigned.

In addition, the CAL program was started in half of the schools in the year 2002-2003, focusing exclusively on children in standard four. The sample was stratified according to gender and language of instruction ("medium", in the official terminology) of the school, and average math test scores in the post-test in the previous year. Table 1 summarizes the allocation of school across different groups in the program. During the school year 2002-2003, 61 schools have a balsakhi in standard three while 62 do not. In standard four, thirty-one schools have a balsakhi and no CAL program, and thirty-one schools have both a balsakhi and the CAL Program. Thirty schools have the CAL program and no extra balsakhi. Thirty-one schools have no balsakhi in standard four (the balsakhi is only in standard three) and no CAL program.

3.2 Sample: Mumbai

To ensure the generalization of the results from the Vadodara study, the balsakhi program in Mumbai was also evaluated in 2001-2002 and 2002-2003. Mumbai was Pratham's birthplace, and Pratham is currently operating various programs throughout the city. We selected one ward (the L-ward) to implement a design similar to the design in Vadodara, including all schools using Hindi, Marathi, and Urdu instruction. In total, 62 schools are included in the study. Schools were stratified according to their scores in a pre-test and the medium of instruction. Half the schools were randomly selected to receive a balsakhi in standard two, and half the schools were randomly selected to receive a balsakhi in standard three. Only standard three children were included in the study. In 2002-2003, schools switched groups, again ensuring that a child who benefited from the balsakhi in 2001-2002 also benefited from it in 2002-2003.

3.3 Outcomes

The main outcome of interest is whether the interventions resulted in any improvement in cognitive skills.

In Vadodara, children were tested in November, 2000, and March, 2001. In 2001, children were tested at the beginning of the school year (August), in November, and in March, 2002. Children were then tested in August, 2002, and November, 2002. In Mumbai, children were tested in October, 2001, March, 2002, and in August, 2002, and will be tested again at the end of February, 2003.

In Vadodara, the same test is used for standard three and four children, so that the scores can be directly compared across grades. Scores on the pre- and post-tests can also be directly compared, as the format of the questions and the competencies tested remain the same. The exam comprises two parts: A math section and a language section. In Vadodara, both parts focused on competencies that the Vadodara Municipal Corporation (VMC) prescribe for children in standards one through four. On the math exam, for example, tasks ranged from basic number recognition, counting, and ordering of single digit numbers to ordering of two digit numbers, addition of single and two digit numbers, and basic word problems. In Mumbai, the test focused on competencies for standards one through three.

The first year of the program (2000-2001) allowed Pratham to make significant progress in developing a testing instrument (the initial test was too difficult) and effective testing procedures to prevent cheating and exam anxiety. The test is administered in both cities by Pratham, with the authorization of the municipal corporation. At least three people are present in the classroom during the test, to minimize cheating.[4] To minimize attrition, Pratham returns to the schools multiple times and children who still failed to appear and who could be tracked down were administered a make-up test outside of school.

In Vadodara, the school year 2001-2002 was disturbed by massive inter-communal riots in Vadodara and Amhedabad, sparked by an attack on a train of Hindu workers. Although a post-test was conducted in March (after the riots had receded), attrition was high. This preliminary draft thus focuses on four set of tests: A pre-test in August, 2001, a post-test in March, 2002, a second pre-test in August, 2002, and a second post-test in November 2002.[5]

Another outcome of interest is attendance and school dropout rates, which are collected weekly by Pratham employees (to avoid using the official rolls, which are often manipulated).[6]

[4] In Mumbai, since administration of the pre-test was less than satisfying at the first attempt, we conducted a second pre-test, which we use as the basis for the analysis.

[5] The results of the first year of the program do not significantly change if we use the mid- or post-test: There was no further improvement (or deterioration) of the performance in the treatment schools relative to the control schools between the mid-test and the post-test.

[6] Attendance data is not analyzed in this report: In 2001-2002, teachers in some schools often refused to let the research assistants count the number of children present, resulting in biased data. We are currently working to differentiate the schools where this occurred, and to clean the data collected in 2002-2003.

Finally, we collected several intermediate outcomes for the CAL program. In particular, at every session, we collected data on the games played by the children, and at what level they were played. This data will be analyzed in the next draft of this paper.

3.4 Statistical Framework

3.4.1 Effect of the Balsakhi Program

Given the randomized allocation of both programs, we expected the results of the 2001 pre-test to be similar between the treatment and control groups. The results of the 2002 pre-test may be different in the treatment and control schools in standard four in Vadodara, and in standard three in Mumbai, since they may reflect long-lasting benefits of the previous year's program, for the children who were in the same school in the previous year. In Vadodara, the cross-cutting design (where in each school children from one standard are in the comparison group, and children from the other standard are in the treatment group) ensures that even if a "good school" was picked up in the treatment group for a standard, it figures in the comparison group of the other standard, ensuring that the averages across the standard are likely to be very similar.

Noting y_{iktj} the test score of child i in grade g in school j in test k (k can be "pre" or "mid" in Vadodara, or "pre" and "post" in Mumbai) in year t (year 1 of the experiment, or year 2 of the experiment), we start by comparing test scores in the pre-test and the post-test, in each city and standard, and we run the following regression to assess the significance of the difference:

$$y_{igjtPRE} = \alpha + \beta B_{jgt} + \upsilon_{jgPREt} + \epsilon_{igjPREt}, \qquad (1)$$

where B_{jg} is a dummy indicating whether school j receives the balsakhi in year t in standard g.

This regression is run separately in each standard, year and city, and combining standards three and four in Vadodara. It is run separately for the math exam, the verbal exam, and the total score on the exam. The standard errors are clustered at the school level.[7]

We then run the same regression in the post-period ($k = POST$):

[7] If instead we used a random effect model, with a nested random effect at the school and division level, the point estimates are very similar, and the estimated standard errors are smaller, making the results more significant.

$$y_{igjPOSTt} = \alpha + \beta B_{jg} + u_{jg} + \epsilon_{igjPOSTt}. \quad (2)$$

This provides a first estimate of the program effect. Because tests scores are very strongly auto-correlated, the precision of the estimate is increased by relying on differences-in-differences estimates. This estimate also controls for any pre-existing differences between the treatment and the control group.

$$y_{igjkt} = \lambda + \delta B_{jg} + \gamma B_{ig} * POST + u_{jg} + \epsilon_{igjkt}. \quad (3)$$

In the absence of large differences between the treatment and the control groups, the coefficients β in equation 2 and in equation 3 should be similar.

3.4.2 Effect of the CAL Program, and Interaction

The pre-program differences and the effect of the CAL program can be estimated with equations similar to 1 and 2, with the treatment dummy for the CAL program replacing the treatment dummy for the Balsakhi program.

However, since the Balsakhi (B_{igt}) and the CAL (C_{igt}) program were run in the same schools, we can estimate the effects of each program together, in a single regression:

$$y_{igjkt} = \lambda + \delta_1 B_{jg} + \delta_2 C_{jg} + \gamma_1 B_{ig} * POST + \gamma_2 C_{ig} * POST + u_{jg} + \epsilon_{igjkt}. \quad (4)$$

This equation provides estimates of the average effect of each program, controlling for receiving the other one. Because the probability of receiving the CAL program conditional on receiving the balsakhi program is, by construction, the same as the probability of receiving the CAL program conditional on NOT receiving the balsakhi program, but the estimates of each program will be more precise, since the regression controls for the “noise” introduced by the other program.

Finally, we can examine whether the CAL program and the Balsakhi program have interaction effects, using the following specification:

$$y_{igjkt} = \lambda + \delta_1 B_{jg} + \delta_2 C_{jg} + \gamma_3 B_{ig} * POST + \gamma_4 C_{ig} * POST + \gamma_5 B_{ig} * C_{ig} * POST + + u_{jg} + \epsilon_{igjkt}. \quad (5)$$

In this regression, $\hat{\gamma_3}$ is an estimate of the effect of the balsakhi program in schools where there is no CAL program, $\hat{\gamma_4}$ is an estimate of the effect of the CAL program in schools where there is no balsakhi program, and $\hat{\gamma_5}$ is an estimate of the difference between the effect of the CAL program in schools that have a balsakhi and in schools that do not have a balsakhi.[8]

Finally, we estimate the effect of the CAL program (using equation 4 in the two sub-samples with and without balsakhi), and the effect of the Balsakhi program in the sample with and without the CAL program.

4 Results

4.1 Descriptive Statistics: Level of Competencies and Pre-intervention Differences

Tables 2 through 5 present the descriptive statistics of the test scores for all samples used in this analysis (year 1 and 2 in Vadodara, and year 1 in Mumbai). The scores are presented both as raw scores (out of 50 points for each test in Vadodara, and 40 points for Mumbai) and normalized relative to the distribution of the pre-test score in the Control group in each city and year.[9]

The randomization appears to have been successful: Neither in Mumbai nor in Vadodara are there any large systematic differences between the pre-test score and the post-test score. In the first year in Vadodara, children in the treatment group perform slightly worse than those in the control group in standard four. In year 2, children in treatment group perform slightly better in standard 4, which may partly reflect the effect of the previous year's program (although the short school year and the population movements due to the riots make it unlikely that the program would have had persistent effects). None of these differences, however, are significant.

The raw scores give an idea of how little these children actually know. In standard three in Vadorara, for example, the average math score is about 8, both in the control and treatment groups. Since one math question provides students answers from which they could guess, on

[8] Or, alternatively, the difference between the effect of the balsakhi program in schools that have or do not have the CAL program.

[9] We subtract the mean of the control group in the pre-test, and divide by the standard deviation.

average a student who knows nothing will score 0.9 points. If a student can just consistently order two numbers and add two single digit numbers, she can score the 7 additional points needed to match the average third standard performance. Children in the bottom in the distribution of pre-tests score on average between 5.2 and 5.5 points on the two tests combined, indicating that they do not know how to count and have trouble even copying sentences, a task that requires no comprehension of the actual words.

4.2 Attrition

Table 6 presents the attrition that occurred between the pre-test and the mid-test or post-test in both years in Vadodara, and in year 1 in Mumbai, broken down by treatment status. Attrition was higher in Vadodara in year 1 than in both Vadodara in year 2 and in Mumbai, which is due in part to the tense communal situation. The post-test was run after the riots, but the team was able to track down some of the children who did not appear for the exam. Encouragingly, attrition rates are not higher in the comparison group than in the treatment group: In year 1 in Vadodara, attrition was 19% in the balsakhi treatment group, and 19% in the comparison group. In year 2, attrition was 9.3% in the balsakhi treatment group, and 8.5% in the balsakhi comparison group. In Mumbai, attrition was 6.9% in the treatment group, and 7.3% in the comparison group. Likewise, attrition is similar in the CAL comparison and treatment groups.

The fact that there was no differential attrition rate in the treatment and control groups suggests that the estimate of the treatment effects should not be biased, unless different types of people drop out from the sample in the treatment and the control groups (Angrist, 1995; and Powell, 199X). This does not seem to be the case here: The second row in each panel presents the difference between the score at the pre-test of children who were not present at the post-test, by treatment status. The third and sixth columns present the differences-in-differences in the treatment and comparison groups. Children who will eventually leave the sample tend to be at the bottom of the distribution of the pre-test scores. However, the difference is very similar in the treatment and control groups, in both years and for both programs. The only difference is in Mumbai, where the attritors seem to have been performing relatively better in the treatment group (although the difference is not significant), which could come from the fact that the program encouraged the weaker children to stay in school, making it easier to track them down subsequently. If anything, this would bias the estimation results downward.

Finally, both the attrition and the difference in test scores are also similar among the bottom 20 children in each school, those who were the most likely to be assigned to a balsakhi.

4.3 Effects of the Balsakhi Program

Tables 3 and 4 present the first estimates of the effect of the balsakhi program as simple differences between the post-test scores in the treatment and control groups. In all years and standards, for both tests, and in both cities, and for all subgroups, the difference between treatment and control groups are positive. In the first year in Vadodara, the differences in post-test between treatment and control groups (for both tests combined) was 0.17 in standard three, and 0.11 in standard four. Among the bottom third of the children, it was 0.25 in standard three, and 0.15 in standard four. These differences are significant only for the bottom third of the children. The results in Mumbai are remarkably similar, with the overall test score improving by 0.14 standard deviation, and that of the bottom third of the children by 0.24 standard deviation. In the second year of the program in Vadodara, the effects are much larger: The difference in test scores is 0.465 in standard three, and 0.431 in standard four. Among children who were at the bottom of the distribution of the pre-test scores, the differences between treatment and control groups in the mid-test is 0.6 standard deviations in both standards. This is a very large difference, by the standard of most interventions in the education literature. In year two, all of the differences between treatment and control groups are significant.

Because test scores have a strong persistent component, the precision of these estimates can be improved significantly, however, by turning to a differences-in-differences specification (equation 3). Since the randomization appeared to be successful, and almost all the children who took the post-test also took the pre-test, the point estimates should be similar in the simple differences and the differences-in-differences specification. The confidence intervals should, however, be tighter. Table 7 presents differences-in-differences estimates of the effects of the balsakhi program, in various years, cities, standards, and sub-groups. Overall, the balsakhi program increased tests scores in both years. In year 1 in Vadorara, the balsakhi program increased the average test scores by 0.20 standard deviation in math, 0.11 in language, and 0.164 overall (column 6). All of these estimates are significant. The differences-in-differences results are indeed very similar to the corresponding simple differences results. In Mumbai, it increased the test scores by 0.20 in math, 0.07 in language, and 0.095 overall (only the math increase is significant). Strikingly, the results are very similar in Mumbai and in Vadodara, despite the very different settings where these interventions took place. The results are also very similar in standard three and four in Vadodara (for example, the overall increase in test scores is 0.164 in standard three, and 0.160 in standard four).

We then separate the children into two groups, those who would have been likely to be assigned to a balsakhi if there was one in their class (the bottom 20 children in each grade in each school, on the basis of their pre-test distribution), and the rest of the children. The effects of the balsakhi program is stronger among the bottom 20 children: The increase in the overall test score is 0.30 in Mumbai, and 0.24 in Vadodara, while the other children's scores do not increase in Mumbai, and increase by 0.11 in Vadodara (the improvement in Vadodara is significant). The fact that, at least in Vadodara, children who did not work with the balsakhi improve their test scores is not surprising, since the balsakhi removes the weakest children from the classroom. But the bottom of the class definitely benefits more from the balsakhi program: This suggest that the balsakhi is actually doing something useful with the children she pulls out. Overall, the program not only increases the overall test scores, but reduces inequality in achievement: To show this, we separate the children into three groups based on the distribution of the pre-test scores. Uniformly, the treatment effect is the largest for children in the bottom third of the class.

The treatment effects is larger in the second year of the program: The overall effect, pooling standard three and four children together is an improvement of 0.39 standard deviations in the sum of the test scores. The treatment effects are very similar in math (0.38) and language (0.35), and for standard three (0.41) and standard four children (0.36). Here again, the bottom 20 children in each class benefit more than the other (0.54 versus 0.30), and the treatment effects increase at the bottom of the distribution (it is 0.59 in the bottom third, 0.31 in the middle third, and 0.28 in the top third). All of the treatment effects, however, are strongly significant. These results remain very similar to the simple differences results.

In this table, we also present the treatment effects distinguished by boys and girls: Across the board, the effects are very similar across gender.

4.4 Effect of the Computer Assisted Learning Program

Table 5 displays the results of estimating equations 1 and 2, with the CAL program as the treatment dummy. There was a small and insignificant negative effect of the pre-test verbal and math scores. In the post-test score, there is a very small insignificant positive difference between the treatment and control groups in math, and a small negative difference in the verbal score.

Table 8 display the results of estimating equation 4 and 5. Consistent with the previous results, the balsakhi program improves the average math test scores by 0.35 standard deviation, while the CAL program improves them by an insignificant 0.10 standard deviation. The Balsakhi also improves the verbal tests scores (by 0.32 standard deviations), while the CAL program has an insignificant negative effect on the verbal test scores. On average, the CAL program appears to be ineffective, while the balsakhi program improves overall test scores by 0.36 standard deviations. In all specifications, we can reject the hypothesis that the CAL program is as effective as the balsakhi program.

The CAL program and the balsakhi do seem to interact positively (column 2), though the coefficient of the interaction remains insignificant. The CAL program has the biggest positive effect on the math score in the balsakhi sub-sample: The coefficient is 0.16, with a t-statistic of 1.68.

Computing the math score only for questions directly targeted by the software we used produces estimates that are of the same magnitude, but more precise (there might be more randomness in the remaining questions). The CAL program shows a significant effect in the balsakhi group: In this group, it increases the score on the appropriate math questions by 0.18 standard deviation (significant at 95%) (Table 8, panel B).

Table 9 presents the results of 4 and 5 in the middle third, the bottom third, and the top third of the distribution of total pre-test scores. As we saw before, the balsakhi program is most effective in the bottom third of the class. The effectiveness of the CAL program, however, seems to be independent of the initial score: The estimate varies between 0.07 and 0.1 standard deviations, and is never significant.[10]

[10] We also examined the effect in the top quintile of the pre-test distribution, to test the hypothesis that only children with sufficient mastery of the material really benefited from the program. The results for the top quintile are not any more encouraging than the results of the top third.

We are now in the process of analyzing the data we have and collecting new data to understand why the CAL program was not as effective as one might have hoped: Do children actually use the software so that they can learn, or do they just randomly play the games without trying to understand the questions posed by the games? Do they fail to realize that an addition on paper is the same as an addition done on a computer? The CAL program is also scheduled to continue at least for another semester, and probably for another year, which will provide insights as to whether this is a temporary adjustment problem or whether this year's results are indicative of what is going to happen in the longer run.

5 Cost Benefit Analysis

We now use these estimates and data on the cost of the programs to evaluate the relative effectiveness of three possible options: Hiring additional teachers, hiring balsakhi instructors, and continuing the balsakhi program. Table 11 shows the cost per student per year of the balsakhi program (248 rupees, or 5 dollars),[11] the recurring expenditures of the CAL program (784 rupees), the cost of the CAL program including the cost of the computers, assuming they are depreciated over five years (1,289 rupees), and the cost of having both the balsakhi and the CAL program in the same school. Finally, we present teacher salary cost of the Vadodara Municipal Corporation (3,168 rupees).

Table 12 combines these numbers with the test score improvements over the pre- to mid/post-year period in years 1 and 2. Note that we use the yearly cost and the improvement in test scores over only half a year. The comparison across programs remain valid. Since the mid/post-test was using exactly the same structure as the pre-test, improvement of the control children between the pre- and the mid-test provides a measure of the effect of being in school for four months, which can be compared with the effect of having a balsakhi over the same period. Clearly, this is only suggestive, since things other than being in school may have happened to the children over the time period. In year 1, the improvement between the pre-test and the mid-test was a bit more than twice the estimate of the treatment effect. The ratio of the cost, however, is 12. This calculation suggests that the average balsakhi is 4.5 times more cost effective (in terms of

[11] The denominator includes all students in standards three and four in the treatment schools, since we also will use average test scores.

improvement in test scores) than the average teacher. In the second year, the balsakhi program appears to be 6.7 times more cost effective than the average teacher. It is important to note that this results *do not* suggest that teachers should be replaced by balsakhis, since balsakhis are always complementing the teachers. It provides some evidence that if the Vadodara Municipality wanted to spend additional resources, hiring balsakhis may be a more effective way to do it than hiring additional teachers. The CAL program seems to be less cost effective than the teachers. A combination of the CAL program and the balsakhi program is slightly more cost effective than a teacher.

6 Conclusion

This paper reports the preliminary results of two interventions: A remedial education program and a computer assisted learning program. The remedial education program has already shown that it can be brought to scale, since it is already reaching tens of thousands of children across India. Evaluations conducted in two cities over two years suggest that this is a remarkably effective and cost effective program: Test scores of children who benefited from the program improved by 0.16 standard deviations in the first year, and 0.39 standard deviations in the second year. At the margin, the program is 4.5 to 6.7 times more effective than resources spent on teachers. Results are even stronger for children in the bottom of the distribution (in the bottom third of the distribution, the program improved tests score by 0.22 standard deviations in the first year, and 0.58 in the second year).

The computer assisted learning program, appears to be less effective and even less cost effective than hiring new teachers. It is a bit troubling in the context of the investment in similar programs that are taking place all over India. The results, however, should be considered as preliminary, as the program had only been in operation for four months when the children were tested.

Table 1: Vadodara Sample Design

Standard	Study Group	Number of Schools
Three	Balsakhi	61
	No Balsakhi	62
Four	Balsakhi + CAL	31
	Balsakhi	31
	CAL	30
	No Program	31

Table 2: Summary Statistics: Vadodara, Year 1

	PRE Treatment	PRE Control	PRE Difference	MID Treatment	MID Control	MID Difference
STANDARD 3, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	2596	2527	69	2094	2069	69
AVERAGE SCORE (POINTS) Math	13.125	13.216	-0.091	17.714	15.759	1.955
			(.962)			(1.154)
Verbal	11.646	11.075	0.571	19.264	17.785	1.478
			(.871)			(1.035)
Total	24.771	24.291	0.480	36.978	33.545	3.433
			(1.748)			(2.123)
AVEAGE SCORE (NORMALIZED) Math	-0.008	0.000	-0.008	0.391	0.221	0.170
			(.084)			(.1)
Verbal	0.059	0.000	0.059	0.840	0.688	0.152
			(.089)			(.106)
Total	0.024	0.000	0.024	0.632	0.461	0.171
			(.087)			(.106)
STANDARD 4, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	2414	2619	-205	1960	2178	-205
AVERAGE SCORE (POINTS) Math	22.074	22.608	-0.534	25.276	23.628	1.647
			(.93)			(1.094)
Verbal	17.005	17.635	-0.631	24.899	24.258	0.641
			(.839)			(1.158)
Total	39.078	40.243	-1.165	50.174	47.886	2.288
			(1.711)			(2.198)
AVEAGE SCORE (NORMALIZED) Math	-0.046	0.000	-0.046	0.231	0.088	0.142
			(.08)			(.095)
Verbal	-0.059	0.000	-0.059	0.677	0.617	0.060
			(.078)			(.108)
Total	-0.055	0.000	-0.055	0.472	0.364	0.109
			(.081)			(.105)
STANDARD 3, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	848	854	-6	663	658	-6
AVERAGE TOTAL SCORE (POINTS)	5.539	5.178	0.361	21.837	16.847	4.991
			(.26)			(1.807)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.934	-0.952	0.018	-0.122	-0.371	0.249
			(.013)			(.09)
STANDARD 4, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	847	834	13	668	672	13
AVERAGE TOTAL SCORE (POINTS)	16.566	16.466	0.099	31.540	28.397	3.143
			(.504)			(1.834)
AVEAGE TOTAL SCORE (NORMALIZED)	-1.126	-1.131	0.005	-0.414	-0.563	0.149
			(.024)			(.087)
STANDARD 3, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	886	823	63	719	682	63
AVERAGE TOTAL SCORE (POINTS)	19.713	19.371	0.343	33.471	30.364	3.108
			(.283)			(2.011)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.228	-0.245	0.017	0.457	0.303	0.155
			(.014)			(.1)
STANDARD 4, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	785	901	-116	633	750	-116
AVERAGE TOTAL SCORE (POINTS)	38.093	38.437	-0.344	49.517	45.895	3.622
			(.367)			(1.975)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.102	-0.086	-0.016	0.441	0.269	0.172
			(.017)			(.094)
STANDARD 3, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	862	850	12	712	729	12
AVERAGE TOTAL SCORE (POINTS)	48.890	48.258	0.632	54.618	51.593	3.025
			(1.081)			(2.504)
AVEAGE TOTAL SCORE (NORMALIZED)	1.225	1.194	0.031	1.511	1.360	0.151
			(.054)			(.125)
STANDARD 4, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	782	884	-102	659	756	-102
AVERAGE TOTAL SCORE (POINTS)	64.451	64.516	-0.064	69.695	67.185	2.510
			(.754)			(2.221)
AVEAGE TOTAL SCORE (NORMALIZED)	1.151	1.154	-0.003	1.401	1.281	0.119
			(.036)			(.106)

Table 3: Summary Statistics: Vadodara Year 2

	PRE Treatment	Control	Difference	MID Treatment	Control	Difference
STANDARD 3, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	3108	2884	224	2819	2639	224
AVERAGE SCORE (POINTS) Math	8.377	8.033	0.344	20.969	16.799	4.170
			(.692)			(1.103)
Verbal	11.067	10.874	0.192	23.162	19.622	3.540
			(.693)			(.899)
Total	19.444	18.907	0.537	44.131	36.421	7.711
			(1.299)			(1.928)
AVEAGE SCORE (NORMALIZED) Math	0.037	0.000	0.037	1.380	0.935	0.445
			(.074)			(.118)
Verbal	0.022	0.000	0.022	1.431	1.019	0.412
			(.081)			(.105)
Total	0.032	0.000	0.032	1.523	1.057	0.465
			(.078)			(.116)
STANDARD 4, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	3183	3206	-23	2911	2940	-23
AVERAGE SCORE (POINTS) Math	15.444	14.850	0.593	26.570	21.737	4.832
			(.888)			(.798)
Verbal	17.478	16.620	0.859	27.999	23.832	4.167
			(.832)			(.821)
Total	32.922	31.470	1.452	54.569	45.569	8.999
			(1.643)			(1.547)
AVEAGE SCORE (NORMALIZED) Math	0.045	-0.005	0.050	0.982	0.575	0.407
			(.075)			(.067)
Verbal	0.078	-0.005	0.083	1.098	0.694	0.404
			(.081)			(.08)
Total	0.064	-0.006	0.070	1.100	0.669	0.431
			(.079)			(.074)
STANDARD 3, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1006	994	12	891	908	12
AVERAGE TOTAL SCORE (POINTS)	3.844	3.587	0.257	34.453	24.523	9.930
			(.168)			(2.166)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.909	-0.925	0.016	0.938	0.339	0.599
			(.01)			(.131)
STANDARD 4, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1045	1164	-119	931	1056	-119
AVERAGE TOTAL SCORE (POINTS)	11.481	11.560	-0.079	41.064	29.077	11.988
			(.32)			(1.38)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.962	-0.959	-0.004	0.454	-0.120	0.574
			(.015)			(.066)
STANDARD 3, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1044	912	132	960	838	132
AVERAGE TOTAL SCORE (POINTS)	14.814	14.831	-0.017	41.580	35.356	6.225
			(.184)			(1.904)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.247	-0.246	-0.001	1.369	0.993	0.376
			(.011)			(.115)
STANDARD 4, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1015	1041	-26	944	964	-26
AVERAGE TOTAL SCORE (POINTS)	29.644	29.572	0.073	52.174	45.338	6.836
			(.276)			(1.422)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.093	-0.097	0.003	0.985	0.658	0.327
			(.013)			(.068)
STANDARD 3, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1058	978	80	968	893	80
AVERAGE TOTAL SCORE (POINTS)	38.846	38.280	0.566	55.569	49.517	6.052
			(.968)			(1.886)
AVEAGE TOTAL SCORE (NORMALIZED)	1.204	1.170	0.034	2.213	1.848	0.365
			(.058)			(.114)
STANDARD 4, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1123	1001	122	1036	920	122
AVERAGE TOTAL SCORE (POINTS)	55.836	56.596	-0.760	68.887	64.742	4.145
			(.877)			(1.464)
AVEAGE TOTAL SCORE (NORMALIZED)	1.161	1.197	-0.036	1.785	1.587	0.198
			(.042)			(.07)

Table 4: Summary statistics, Bombay, year 1

	PRE			MID		
	Treatment	Control	Difference	Treatment	Control	Difference
STANDARD 3, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	933	807	126	859	737	126
AVERAGE SCORE (POINTS) Math	13.818	14.370	-0.552	17.065	16.038	1.027
			(.894)			(1.057)
Verbal	17.867	17.240	0.627	20.034	18.839	1.195
			(.861)			(.804)
Total	31.969	31.101	0.868	36.945	34.919	2.027
			(1.744)			(1.964)
AVEAGE SCORE (NORMALIZED) Math	-0.068	0.000	-0.068	0.333	0.206	0.127
			(.11)			(.131)
Verbal	0.076	0.000	0.076	0.340	0.195	0.146
			(.105)			(.098)
Total	0.058	0.000	0.058	0.392	0.256	0.136
			(.117)			(.132)
STANDARD 3, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	321	279	42	287	259	42
AVERAGE TOTAL SCORE (POINTS)	15.179	14.567	0.613	24.217	20.605	3.612
			(.617)			(1.963)
AVEAGE TOTAL SCORE (NORMALIZED)	-1.067	-1.108	0.041	-0.461	-0.703	0.242
			(.041)			(.132)
STANDARD 3, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	304	229	75	280	225	75
AVERAGE TOTAL SCORE (POINTS)	31.925	32.223	-0.298	37.869	35.175	2.695
			(.372)			(1.384)
AVEAGE TOTAL SCORE (NORMALIZED)	0.055	0.075	-0.020	0.454	0.273	0.181
			(.025)			(.093)
STANDARD 3, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	308	299	9	292	253	9
AVERAGE TOTAL SCORE (POINTS)	48.316	47.844	0.472	48.508	48.122	0.386
			(.456)			(1.203)
AVEAGE TOTAL SCORE (NORMALIZED)	1.154	1.122	0.032	1.167	1.141	0.026
			(.031)			(.081)

Table 5: Summary Statistics: CAL Program (Vadodara, year 2)

	PRE			MID		
	Treatment	Control	Difference	Treatment	Control	Difference
STANDARD 4, ALL						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	3116	3273	-157	2888	2963	-157
AVERAGE SCORE (POINTS) Math	14.744	15.529	-0.785	24.370	23.919	0.451
			(.886)			(.914)
Verbal	16.919	17.170	-0.251	25.729	26.077	-0.348
			(.835)			(.915)
Total	31.663	32.698	-1.035	50.099	49.996	0.103
			(1.643)			(1.768)
AVEAGE SCORE (NORMALIZED) Math	-0.014	0.052	-0.066	0.797	0.759	0.038
			0.075			0.077
Verbal	0.024	0.048	-0.024	0.878	0.912	-0.034
			(.081)			(.089)
Total	0.004	0.053	-0.050	0.886	0.881	0.005
			(.079)			(.085)
STANDARD 4, BOT THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1108	1101	7	1008	979	7
AVERAGE TOTAL SCORE (POINTS)	11.630	11.415	0.215	35.615	33.745	1.870
			(.324)			(1.876)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.955	-0.966	0.010	0.193	0.103	0.090
			(.016)			(.09)
STANDARD 4, MID THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1004	1052	-48	944	964	-48
AVERAGE TOTAL SCORE (POINTS)	29.639	29.577	0.062	48.721	48.719	0.003
			(.277)			(1.529)
AVEAGE TOTAL SCORE (NORMALIZED)	-0.093	-0.096	0.003	0.820	0.820	0.000
			(.013)			(.073)
STANDARD 4, TOP THIRD						
NUMBER OF TESTS (AMONG PRE_TEST TAKERS)	1004	1120	-116	936	1020	-116
AVERAGE TOTAL SCORE (POINTS)	55.795	56.553	-0.758	67.087	66.801	0.286
			(.874)			(1.515)
AVEAGE TOTAL SCORE (NORMALIZED)	1.159	1.195	-0.036	1.699	1.685	0.014
			(.042)			(.073)

Table 6: Attrition patterns

	Bombay			Vadodara, year 1			Vadodara, year 2			Vadodara, year 3		
	Balsakhi	No Balskahi	difference	Balsakhi	No Balskahi	difference	Balsakhi	No Balskahi	difference	CAL	No CAL	Difference
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Standard 3, ALL												
Percent attrition	0.069	0.073	-0.005	0.193	0.181	0.012	0.093	0.085	0.008			
Difference in score at pre-test	-0.008	-0.199	0.192	-0.130	-0.222	0.092	-0.119	0.072	-0.191			
attriters-stayers			(.211)			(.093)			(.119)			
Standard 4, ALL												
Percent attrition				0.188	0.168	0.020	0.085	0.083	0.002	0.073	0.095	-0.022
Difference in score at pre-test				-0.178	-0.176	-0.002	-0.137	-0.093	-0.044	-0.157	-0.089	-0.069
attriters-stayers						(.077)			(.118)			(.115)

Table 7: Differences in differences estimate of the impact of the balsakhi program, by city and sample

	Standard 3			Vadodara Standard 4		Vadodara 3 and 4 together	
		Vadodara					
	Bombay	Year 1	Year 2	year 1	year 2	Year 1	year 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. MATH							
All	0.204	0.198	0.395	0.186	0.351	0.196	0.380
	(.088)	(.092)	(.109)	(.074)	(.061)	(.053)	(.074)
Bottom 20	0.254	0.314	0.621	0.234	0.465	0.273	0.548
	(.101)	(.103)	(.137)	(.089)	(.084)	(.065)	(.087)
Not Bottom 20	0.174	0.140	0.264	0.138	0.287	0.147	0.284
	(.122)	(.109)	(.117)	(.09)	(.066)	(.066)	(.081)
Bottom Third	0.225	0.255	0.593	0.194	0.530	0.224	0.568
	(.119)	(.084)	(.136)	(.085)	(.065)	(.058)	(.077)
Middle Third	0.093	0.137	0.370	0.203	0.367	0.186	0.390
	(.108)	(.109)	(.13)	(.093)	(.073)	(.071)	(.091)
Top Third	0.085	0.191	0.237	0.117	0.246	0.156	0.238
	(.083)	(.136)	(.131)	(.098)	(.072)	(.071)	(.083)
Boys	0.198	0.231	0.377	0.166	0.386	0.207	0.376
	(.118)	(.099)	(.126)	(.089)	(.075)	(.062)	(.085)
Girls	0.211	0.168	0.409	0.205	0.319	0.186	0.385
	(.089)	(.113)	(.131)	(.086)	(.071)	(.062)	(.085)
B. VERBAL							
All	0.072	0.101	0.371	0.114	0.320	0.109	0.352
	(.087)	(.09)	(.09)	(.075)	(.061)	(.05)	(.062)
Bottom 20	0.114	0.205	0.500	0.149	0.419	0.176	0.466
	(.081)	(.11)	(.11)	(.086)	(.083)	(.064)	(.082)
Not Bottom 20	0.022	0.047	0.302	0.085	0.263	0.068	0.287
	(.145)	(.102)	(.098)	(.084)	(.065)	(.056)	(.064)
Bottom Third	0.232	0.165	0.480	0.074	0.540	0.120	0.524
	(.099)	(.091)	(.115)	(.088)	(.079)	(.061)	(.083)
Middle Third	-0.040	0.137	0.316	0.139	0.249	0.140	0.296
	(.109)	(.101)	(.102)	(.097)	(.073)	(.06)	(.07)
Top Third	0.006	-0.005	0.346	0.124	0.217	0.060	0.277
	(.098)	(.134)	(.114)	(.095)	(.079)	(.072)	(.073)
Boys	-0.014	0.051	0.442	0.107	0.324	0.084	0.375
	(.107)	(.106)	(.113)	(.092)	(.079)	(.06)	(.074)
Girls	0.146	0.148	0.302	0.118	0.318	0.133	0.330
	(.1)	(.103)	(.099)	(.097)	(.069)	(.063)	(.07)
C. TOTAL SCORE							
All	0.095	0.162	0.416	0.160	0.357	0.164	0.394
	(.086)	(.089)	(.1)	(.072)	(.059)	(.048)	(.071)
Bottom 20	0.301	0.279	0.611	0.205	0.471	0.241	0.547
	(.094)	(.107)	(.126)	(.084)	(.081)	(.062)	(.087)
Not Bottom 20	-0.086	0.103	0.305	0.119	0.293	0.116	0.307
	(.121)	(.102)	(.106)	(.085)	(.064)	(.056)	(.075)
Bottom Third	0.186	0.226	0.584	0.145	0.568	0.185	0.587
	(.125)	(.087)	(.131)	(.084)	(.07)	(.058)	(.083)
Middle Third	0.197	0.145	0.373	0.183	0.332	0.173	0.372
	(.084)	(.1)	(.114)	(.091)	(.066)	(.059)	(.083)
Top Third	-0.011	0.107	0.313	0.128	0.247	0.118	0.276
	(.081)	(.13)	(.118)	(.096)	(.074)	(.064)	(.077)
Boys	0.018	0.157	0.442	0.146	0.379	0.158	0.403
	(.101)	(.1)	(.119)	(.088)	(.076)	(.058)	(.083)
Girls	0.160	0.168	0.388	0.173	0.338	0.170	0.386
	(.097)	(.105)	(.116)	(.089)	(.065)	(.058)	(.08)

Table 8 Balsakhi and CAL program, standard 4, all children

	All		CAL		Balsakhi	
			No	Yes	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
A. MATH						
Balsakhi	0.35	0.29	0.29	0.41		
	(.06)	(.089)	(.089)	(.081)		
CAL	0.10	0.04			0.04	0.16
	(.06)	(.073)			(.073)	(.096)
BOTH		0.12				
		(.12)				
	7.83	7.93				
	(.006)	(.005)				
Observations	11702	5926	5926	5776	5880	5822
B. MATH (Only skills covered by computer instruction)						
Balsakhi	0.27	0.20	0.20	0.35		
	(.055)	(.078)	(.079)	(.077)		
CAL	0.10	0.02			0.02	0.18
	(.055)	(.068)			(.068)	(.086)
BOTH		0.16				
		(.11)				
	4.67	4.78				
	(.032)	(.03)				
Observations	11702	5926		5776	5880	
C. VERBAL						
Balsakhi	0.32	0.25	0.25	0.39		
	(.06)	(.08)	(.081)	(.089)		
CAL	-0.02	-0.09			-0.09	0.05
	(.06)	(.08)			(.08)	(.089)
BOTH		0.14				
		(.119)				
	15.97	16.04				
	(.00009)	(.00009)				
Observations	11702	11702	5926	5776	5880	5822
D. BOTH						
Balsakhi	0.36	0.29	0.29	0.43		
	(.058)	(.082)	(.082)	(.083)		
CAL	0.05	-0.02			-0.02	0.12
	(.058)	(.073)			(.073)	(.091)
BOTH		0.14				
		(.116)				
	12.92	13.06				
	(.00042)	(.00039)				
Observations	11702	11702	5926	5776	5880	5822

Table 9: Balskahi and CAL program, standard 4, by pre-test quantile

	Bottom third		Middle third		Top thid	
	(1)	(2)	(3)	(4)	(5)	(6)
A. MATH						
Balsakhi	0.53	0.44	0.37	0.26	0.25	0.31
	(.071)	(.1)	(.072)	(.105)	(.071)	(.098)
CAL	0.10	0.02	0.07	-0.04	0.09	0.16
	(.069)	(.074)	(.072)	(.086)	(.07)	(.104)
BOTH		0.16		0.22		-0.13
		(.141)		(.143)		(.141)
	19.49	18.14	7.44	7.60	2.14	1.94
	(.00002)	(.00003)	(.007)	(.006)	(.145)	(.166)
Observations	3974	1958	3816	1928	3912	2040
B. VERBAL						
Balsakhi	0.54	0.37	0.25	0.20	0.22	0.24
	(.076)	(.106)	(.07)	(.098)	(.078)	(.096)
CAL	-0.02	-0.17	-0.10	-0.14	0.02	0.05
	(.075)	(.085)	(.071)	(.105)	(.078)	(.123)
BOTH		0.33		0.09		-0.05
		(.15)		(.141)		(.157)
	32.36	30.16	11.88	11.85	2.87	2.68
	(.)	(.)	(.0007)	(.00071)	(.09185)	(.10342)
Observations	3974	3974	3816	3816	3912	3912
BOTH						
Balsakhi	0.57	0.44	0.33	0.25	0.25	0.30
	(.072)	(.098)	(.065)	(.092)	(.072)	(.092)
CAL	0.05	-0.07	-0.01	-0.09	0.06	0.11
	(.07)	(.073)	(.065)	(.088)	(.071)	(.111)
BOTH		0.25		0.17		-0.10
		(.142)		(.129)		(.144)
	29.43	27.26	11.63	11.73	2.84	2.60
	(.0000002)	(.0000005)	(.00079)	(.00075)	(.09379)	(.10881)
Observations	3974	3974	3816	3816	3912	3912

Table 10: Balsakhi and CAL program, standard 4, by gender

	All		Subsample		Subsample	
			No CAL	CAL	No Balsakhi	Balsakhi
	(1)	(2)	(3)	(4)	(5)	(6)
BOYS						
A. MATH						
Balsakhi	0.39	0.34	0.34	0.44		
	(.074)	(.103)	(.103)	(.107)		
CAL	0.12	0.07			0.07	0.17
	(.075)	(.094)			(.095)	(.115)
BOTH		0.10				
		(.148)				
Observations	5828	5828	2896	2932	2778	3050
B. VERBAL						
Balsakhi	0.16	0.12	0.12	0.20		
	(.039)	(.054)	(.054)	(.054)		
CAL	0.01	-0.04			-0.04	0.05
	(.039)	(.047)			(.047)	(.061)
BOTH		0.09				
		(.077)				
Observations	5828	5828	2896	2932	2778	3050
BOTH						
Balsakhi	0.38	0.31	0.31	0.45		
	(.075)	(.105)	(.105)	(.108)		
CAL	0.07	0.00			0.00	0.14
	(.076)	(.092)			(.092)	(.119)
BOTH		0.14				
		(.15)				
Observations	5828	5828	2896	2932	2778	3050
GIRLS						
A. MATH						
Balsakhi	0.31	0.25	0.25	0.38		
	(.073)	(.114)	(.114)	(.09)		
CAL	0.09	0.03			0.03	0.16
	(.072)	(.087)			(.087)	(.116)
BOTH		0.13				
		(.145)				
Observations	5874	5874	3030	2844	3102	2772
B. VERBAL						
Balsakhi	0.16	0.13	0.13	0.18		
	(.035)	(.047)	(.047)	(.051)		
CAL	-0.03	-0.05			-0.05	0.00
	(.035)	(.051)			(.052)	(.046)
BOTH		0.05				
		(.069)				
Observations	5874	5874	3030	2844	3102	2772
BOTH						
Balsakhi	0.34	0.28	0.28	0.40		
	(.068)	(.099)	(.099)	(.09)		
CAL	0.03	-0.03			-0.03	0.09
	(.067)	(.088)			(.089)	(.101)
BOTH		0.13				
		(.134)				
Observations	5874	5874	3030	2844	3102	2772

Table 11: Cost comparison

	Cost per year (dollars)	Cost per Year (Rupees)	Students	Rps/student per year
Balsakhi	31556	1420000	5730	248
CAL, w/o capital expenditures	50338	2265200	2888	784
CAL, with capital expenditures	82738	3723200	2888	1289
Balsaki+CAL				1537
Teachers in Vadodara Primary Schools	39224300	172740888	54525	3168

Table 12: Cost benefit analysis

	Improvements in Test scores		Rupees per standard deviation	
	Standard 3	Standard 4	Standard 3	Standard 4
Year 1				
Balsakhi	0.16	0.16	1549	1549
Mid to pre test difference (control)	0.36	0.46	8800	6887
Year 2				
Balsakhi	0.416	0.357	596	694
CAL, w/o capital expenditures		0.1		7843
CAL, with capital expenditures		0.1		12892
Balsaki+CAL		0.45		3416
Mid to pre test difference (control)	1.057	0.675	2997	4693