# Remedying Education: Evidence from Two Randomized Experiments in India

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

November 2004

**Abstract**

This paper presents the results of two overlapping two-year randomized evaluations conducted in Mumbai and Vadodara, India, designed to evaluate ways to improve the quality of education in urban slums. A remedial education program hires young women from the community to teach basic literacy and numeracy skills to children lagging behind in government schools. Children are removed from the regular classroom for half a day. We find the program to be very effective: It increased average test score of all children in treatment schools by 0.14 standard deviations in the first year, and 0.28 in the second year. A computer assisted learning program provided each child in the fourth standard with two hours of shared computer time per week, in which students played educational games that reinforced mathematics skills. The program was also very effective, increasing math scores by 0.36 standard deviation the first year, and 0.54 the second year. Two instrumental variable strategies suggest that the effect of the remedial education program benefited only children who participated. This suggest that reducing class size without changing pedagogy many not be beneficial.

*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: Pratima Bandekar, Rukmini Banerji, Lekha Bhatt, Madhav Chavan, Shekhar Hardikar, and Rajashree Kabare. We thank Jim Berry and Marc Shotland, Mukesh Prajapati and Nandit Bhat for their excellent work coordinating the fieldwork. They were helped by Aditya Naraj and many others. Kartini Shastry provided superb research assistance. For financial support, we thank the ICICI corporation, the World Bank, the Alfred E. Sloan Foundation and the MacArthur Foundation Network on the Costs of Inequality.

# 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 burgeoning 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 (1997) on school uniforms in Kenya, Spohr (1999) on compulsory schooling laws in Taiwan and Vermeersch (2002) on school meals for preschoolers in Kenya. The primary metric by which success is judged in these studies is attendance, and in each of these cases a significant impact was found.

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 gains are 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. By improving school quality, these programs can increase attendance. One ought also 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, one cannot rule out the idea that there is no impact on children's educational achievement *a priori*, because the quality of teaching in many schools leaves much to be desired. A possibility is that providing more of the same (more teachers, more textbooks) would not be effective without a radical change in the way children are taught. 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 the randomized evaluations of two intervention in urban India focused on improving the learning environment in public schools. The interventions is motivated by the belief that children learn very little in school because if they fall behind and feel lost in class, there is no mechanism in the India school system to help them get back on track: they continue to be promoted until the end of the primary school cycle, and teachers continue to teach the curriculum. Both interventions employ different methods to help children learn at their own pace.

The first program, which is run Pratham, a Mumbai-based Non-Governmental Organization, provides remedial education, in small groups, to children that are lagging behind. To keep costs low and ensure a good instructor-student relationship, the program hires young women (the "Balsakhis") who have the equivalent of a high school degree from the local slum communities in which the schools are located. The second program, also run by Pratham, is a computer assisted learning, where children in grade 4 are offered two hours of shared computer time per

[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.

week, during which they are playing games seeking to reinforce their maths competencies.

The evaluation of those two programs offered an opportunity to implement an evaluation design that is often recommended but rarely, if ever, utilized. First, these were two partially overlapping randomized evaluation, with a randomized design. We can therefore be relatively confident of the absence of confounding factors. Second, the programs we study were run on a v large scale (over 15,000 students were included in the study over 3 years). The remedial education program had already clearly demonstrated the ability to scale up in other cities, as the description below will make clear. In other words, there is no risk that what we are evaluating cannot be reproduced elsewhere. Third, we simultaneously carried out randomized evaluations of the remedial education program in two different cities, each of which had its own management team. This reinforces our confidence in the external validity of these results. Finally, we conducted each evaluation over two years, using several tests, making it less likely that the results are a consequence of the newness of the program, or the effect of implementing an evaluation.

Finally, though we find no effect on attendance, we find that both program have a substantial positive effect on children's academic achievement. This is true in all years and cities, despite the instability of the environment (notably a major riot in one of the city in 2003).

The remedial education program increased average test score of all children in treatment schools by 0.14 standard deviations in the first year, and 0.26 in the second year. Moreover, the weaker students, who are the primary target of the program, gained the most. The computer assisted learning increased math scores by 0.36 standard deviation the first year, and 0.51 the second year, and was equally effective for all students.

Moreover, two instrumental variable strategies suggest that the effect of the remedial education program benefited only children who participated. This implies that the effect on the students who actually benefited are very high (0.6 to 1 standard deviation) and that reducing class size without changing pedagogy many not be beneficial. This is in line with the results with the previous literature, which found no impact of increasing resources without affecting the pedagogy.

The results thus suggest that it is possible to dramatically improve the quality of education at a very moderate cost by changing the teaching approach prevalent in most Indian schools.

# 2 The Programs

Pratham was established in Mumbai in 1994, with support from UNICEF, and has since expanded to several other cities in India. Pratham now reaches over 275,000 children in 12 States 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.

## 2.1 Remedial Education: The Balsakhi Program

One of Pratham's core programs is a remedial education program, called the Balsakhi program. This program, in place in many municipal schools, provides a teacher (usually a young woman, recruited from the local community, who has herself finished secondary school) to work with children identified as falling behind their peers. While the exact details vary according to local conditions, the typical instructor meets with a group of approximately 15-20 children in a class for two hours at a time (the school day is about 4 hour long). Instruction focuses on the core competencies the children should have learned in the first and second standards, primarily basic numeracy and literacy skills. 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 program has been implemented in twenty Indian cities, reaching tens of thousands of students. It was started in Mumbai in 1994, and 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 curriculum. Children may feel more comfortable with women from their own communities than teachers, who are often from different backgrounds. As the balsakhi's class size is relatively small, she may tailor the curriculum to the children's 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, and the teacher may be able to focus on more advanced material. Finally, 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 500-750 rupees, or 10-15 dollars, per month) and is easily replicated. There is rapid turnover among the balsakhis (each of them staying for an average of one year, typically 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 balsakhis use whatever space is available (free classrooms, playground, or even hallways when necessary), the program has very low overhead and capital costs.

### 2.2 Computer Assisted Learning

The Computer Assisted Learning (CAL) program takes advantage of both a policy put in place by the government of Gujarat in 2000 as well as the established infrastructure of the balsakhi program. The government delivered four computers to each of 100 municipal 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.

This situation is not isolated. Many in India see Computer Assisted Learning (CAL), as a supplement to regular instruction, as a possible way to improve the quality of education. Good educational software can be reproduced at nominal cost, and well-designed educational games can sustain interest and curiosity even in an otherwise dull school environment. The excitement seems to be particularly strong in India, where the high-tech sector is both successful and visible. Many local governments have started providing computers in schools but without offering much guidance about how the schools should use them. The idea of using computers is particularly attractive in urban public schools and in rural areas where the number of qualified teachers is limited and the quality of existing teachers is notoriously poor. Computers have the potential to both directly improve learning and indirectly increase attendance by making school more attractive.

Unfortunately, despite the general excitement, there exists very little rigorous evidence of the impact of computers on educational outcomes and no reliable evidence for India or other developing countries. Furthermore, what evidence that exists is not particularly encouraging. For example, Angrist and Lavy (2002) evaluate a computer assisted learning program in Israeli schools with disappointing results. Among the fourth and eighth grade students evaluated with math and Hebrew exams, the data show no benefits for computer assisted instruction and provide some evidence that children who received such instruction are actually at a disadvantage. Krueger and Rouse (2003) reports on a randomized evaluation of the language software "Fas for word" commonly used in US classrooms, and find no impact.

It is not clear, however, that these results apply to the use of computers in schools in developing countries since in Israel, the computer-assisted learning replaces time spent in well equipped classrooms with high quality instructors. It is easy to imagine that computers can make a significant improvement in schools in developing countries even if they do not prove to be useful in the developed world.

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 computer 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. In the second year of the program, Pratham teamed up with Media-pro, a compagnie that develop instruction software, to develop a suite of software that more closely followed the curriculum. Children also completed simple worksheets designed to track down their progress at the beginning of each session.

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 comprehension, and when necessary, they helped individual children understand the tasks required of them by the game. All interaction between the students and instructors was driven by the child's use of the various games, and at no time did any of the instructors provide general instruction in mathematics.

Schools where the CAL program was not implemented were free to continue to use the computer at their convenience, but our observation was 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

• Balsakhi

In 2000, when Pratham decided to expand their remedial education (balsakhi) program to cover the entire city of Vadodara, 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 administered an academic test (designed by the Pratham team) to all children in the third standard. They then hired and trained balsakhis, which were sent to half of the schools in Vadodara. Assignment was random, with schools stratified by language ("medium" in the official terminology) 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. This year of the program is best understood as a pilot program.[3]

In July, 2001, the group of schools that had received a balsakhi in the previous 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.

[3]Throughout the paper, we will refer to the year 2001-2002 as "year 1", year 2002-2003 as "year 2" and year 2003-2004 as "year 4."

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 in year 2, standard 4 children in the treatment group benefitted from two years of the balsakhi program. Schools where the balsakhi was assigned in standard four in the year 1 received balsakhi assistance for standard three in year 2. The new schools were randomly assigned to either group with equal probability in the same way that the original schools were assigned. The number of schools and divisions in the two groups are given in Table 1.

In the school year 2003-2004, the program was extended, with some modifications, to 100 of those 122 schools (schools where the instruction was not provided in Gujarati were not included in the program).

• Computer assisted learning

The CAL program was started in approximately half of the municipal primary schools in Vadodara in 2002-2003, focusing exclusively on children in standard four. The sample was stratified according to treatment or control status for the standard four balsakhi program as well as gender, language of instruction of the school, the average math test scores in the post-test in the previous year. Table 1 summarizes the allocation of schools across different groups in the program. In some schools, computers could not physically be installed, either because of space constraints or the lack of electricity to run the computers. These schools are excluded from the comparison as well as the treatment group. Thus, in the final sample for the study, 55 schools received the CAL program and 56 serve as the control group.

The program was continued in 2003-2004, after switching the treatment and comparison groups.

### 3.2 Sample: Mumbai

To ensure the results from the Vadodara study would be generalizable, 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 Gujarati, Hindi, and Marathi schools. In total, 62 schools are included in the study. Schools were stratified according to their scores in a pre-test, as well as by 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. In 2001-2002, data were collected only for standard three children, while in 2002-2003, data were collected for standards three and four. As in Vadodara, children kept their treatment assignment status as the moved from standard two to three (or three to four).

In the second year of the study, the Mumbai program experienced some administrative difficulties. A decision to require balsakhis to pass a competency test resulted in the firing of many balsakhis. Hiring new recruits was complicated by the fact that the administrative staff in L-Ward turned over between year 1 and year 2, and the new staff lacked community contacts necessary for recruitment. Finally, the principals of a couple of schools, hearing that the study was being conducted by a group of Americans, refused balsakhis. Thus, only two thirds of the schools assigned Balsakhis actually received them. (Schools could not refuse testing, because Pratham had obtained written permission for testing from the city administration). Throughout the paper, the schools that were assigned balsakhis but did not get them are included in the treatment group. The analysis then adjust for the fraction of the treatment group that was effectively treated.

### 3.3 Outcomes

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

In the Vadodara pilot year, children were given a pretest in November, 2000, and post-test in March, 2001. In the first full year, the Vadodara pretest was at the beginning of the school year (August 2001), the mid-test was in October 2001, and the post test was in March 2002. In the second full year, children were tested at the beginning of the school year (August 2002), in November 2002, and again in March, 2003. In the first year in Mumbai, children were tested in October, 2001 and March, 2002; in the second year tests were given in August, 2002, and 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-test 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. Tests were similar in Mumbai. In the first year, tests focused on competencies in standards one through three, while in the second year they included standards one through four. In the second year, the same test was used for third and fourth standard children.

The "pilot" 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 was administered in both cities by Pratham, with the authorization of the municipal corporation. At least three Pratham employees were present in the classroom during each 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.

Another outcome of interest is attendance and school dropout rates, which are collected weekly by Pratham employees, who made randomly timed weekly appearances in classrooms to take attendance. (Data from the official rolls was also collected, but administrators have incentives to inflate the attendance data).

Finally, in the second year of the program, in both cities, data were collected on which specific children were sent to the Balsakhi. (Balsakhis work with, on average, about 20 children per school).

### 3.4 Statistical Framework

Given the randomized allocation of both programs, we expect the 2001 pre-test results in the treatment schools to be similar between those in the control. The results of the 2002 pre-test may be different in the treatment and control schools in standard four in Vadodara, as well as

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

standard three and four 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 both cities, the experimental design (in which each school was both in the treatment and comparison group, with one standard in each group) is such that even if a "good school" were in the treatment group for a given standard, the other standard of that "good school" would be in the comparison group, ensuring that the averages across the standard are likely to be very similar.

Denoting $y_{igjk}$ the test score of child $i$ in grade $g$ in school $j$ in test $k$ ($k$ is either "PRE" or "POST"), we start by comparing test scores in the treatment and comparison schools, in each city and standard, and for each type of program (CAL and Balsakhi).

We start by checking that there is no difference between treatment and control schools before the program was run:

$$y_{igjPRE} = \alpha + \beta D_{jg} + \epsilon_{igjPRE}, \tag{1}$$

where $D_{jg}$ is a dummy indicating whether school $j$ is in the treatment group in that particular year in standard $g$, and $\epsilon_{igjPRE}$ the error term.

This regression is run separately in each standard, year and city. 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.[5]

We also implement a bootstrap test of equality of the distribution and first order dominance, to test the hypothesis that the distributions are identical in the treatment and the control groups for the pre-test and the hypothesis that the distribution of the post-test in the treatment group first order stochastically dominates the distribution in the control group. To do this, we use a test proposed by Abadie (2002) which uses the Kolmogorov-Smirnov statistic to measure the discrepancy between the hypothesis of equality of distributions and the data. Denoting $\Gamma_{1,n_1}(y)$ as the empirical distribution function for the treatment outcomes and $\Gamma_{0,n_0}(y)$ as the empirical distribution function for the control group outcomes, the Kolmogorov-Smirnov statistic for two samples is defined as:

[5] If we instead use a nested random effects model (with a classroom effect nested within a school effect), the point estimates are very similar, and the the standard errors are smaller. Clustering is a more conservative approach.

$$T_n^{eq} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y \in \mathbf{R}} |\Gamma_{1,n_1}(y) - \Gamma_{0,n_0}(y)|$$

where $n_1$ is the number of treatment observations and $n_0$ is the number of treatment observations. Modifying this statistic to test the hypothesis that $\Gamma_{1,n_1}(y)$ first-order stochastic dominates $\Gamma_{0,n_0}(y)$, the test statistic we use is

$$T_n^{eq} = \left(\frac{n_1 n_0}{n}\right)^{1/2} \sup_{y \in \mathbf{R}} \left(\Gamma_{1,n_1}(y) - \Gamma_{0,n_0}(y)\right)$$

Since the asymptotic distribution of these test statistics depend on the underlying distribution of the data and are therefore unknown, Abadie proposes a bootstrap strategy to test these hypotheses. We implement a block-bootstrap version of the test to account of the grouped nature of our data (the randomization was performed at the school level).

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

$$y_{igjPOST} = \alpha + \beta D_{jg} + \epsilon_{igjPOST}. \tag{2}$$

This provides a first estimate of the effect of being assigned to the treatment group. We also implement the same test of equality of distribution and first order dominance that are implemented for equation 2.

For all cities and year, except for Mumbai in year 2, the coefficient $\beta$ in equation 2 is also an estimate of the average effect of being a student in a school that was assigned a balsakhi. However, in Mumbai in year 2, because not all schools received a balsakhi (and not all classes within schools where treated), to obtain the average effect of receiving a balsakhi, we use the assignment to the treatment group ($D_{jg}$) as an instrument for whether or not the class of a specific child actually received the balsakhi ($B_{jg}$). In practice, we estimate the following equation:

$$y_{igjPOSTt} = \alpha + \beta B_{jg} + \epsilon_{igjPOST}. \tag{3}$$

The first stage is the equation for whether or not a child's class was actually assigned a balsakhi:

$$B_{jg} = \alpha_1 + \delta_1 D_{jg} + \eta_{igjPOST}. \tag{4}$$

Because tests scores are very strongly auto-correlated, the precision of the estimate is increased by controlling for the child's test score in the pre-test.

We do so in two ways.

First, we control for the pre-test score of child $i$ in equation **??**.

$$y_{igjPOST} = \lambda + \delta B_{jg} + \theta y_{igjPRE}, \tag{5}$$

This specification is in effect a value added specification: it is asking whether children improved more, relative to what they would have been expected to on the basis of their pre-test score, in treatment schools than in comparison schools.

For all years and samples except Mumbai in year 2 $B_{jg} = D_{jg}$, and equation 5 is estimated with OLS. However, for Mumbai in year two (and when both cities are pooled), equation 5 is estimated using instrumental variables, with $D_{jg}$ and $y_{igjPRE}$ used as instruments.

Second, we stack the pre and post data and use the following difference in difference specification:

$$y_{igjk} = \lambda + \delta B_{jg} + \theta POST_k + \gamma(B_{ig} * POST) + \epsilon_{igjk},, \tag{6}$$

where $POST_k$ is a dummy indicating whether the test is the post test. For Mumbai in year two (and when both cities are pooled), equation 6 is estimated with instrumental variables, with $D_{jg}$ (the initial assignment to the treatment group), $POST$, and $D_{jg} * POST$ used as instruments.

We also present an alternative way to estimate the treatment effect in Mumbai, as a specification check. Since every school was supposed to receive a balsakhi in either standard 3 or standard 4, we keep in the sample only the schools that did receive a balsakhi. This means that a school will not be in the comparison school for one standard if the other standard did not receive a balsakhi. In this reduced sample, $B_{jg}$ is equal to $D_{jg}$, and equation 5 and 6 are estimated by OLS. The assumption underlying this specification is that the characteristics that make the school more likely to have a balsakhi have the same influence on the test scores of children in standard 3 and standard 4.

To gain more insight about the impact of the program, we also present estimates of specifications similar to equations 3, 5 and 6 using for $y_{igjk}$ a binary variable indicating whether the child correctly answered the questions indicating competencies for standard 1, 2 and 3, respectively. Finally, we estimate the impact of being in the program for 2 years (for children who were in the treatment group in standard 3 in year 1, and whom the balsakhi has followed in year 2 when they moved to standard 4), by estimating equation 5 and 6 using the pre-test of year 1 as the pre-test, and the post test of year 2 as the post test.

Finally, we study whether there are interactions between the Balsakhi and the Computer Assisted Learning programs in the year where they were run at the same time, by running the regressions:

$$y_{igjPOST} = \lambda + \delta^b D^b_{jg} + \delta^c D^c_{jg} + \mu D^b_{jg} * D^c_{jg} + \theta y_{igjPRE}, \tag{7}$$

where $D^b_{jg}$ indicate Balsakhi treatment and $D^b_{jg}$ indicate computer assisted learning treatment and

$$y_{igjk} = \lambda + \delta^b D^b_{jg} + \delta^c D^c_{jg} + \theta POST_k + \gamma^b D^b_{jg} * POST + \gamma^c D^c_{jg} * POST + \theta D^b_{jg} * D^c_{jg} * POST + \epsilon_{igjk}, \tag{8}$$

# 4 Results: Pre-intervention difference and attrition patterns

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

Tables 2 through 4 present the descriptive statistics of the test scores for all samples used in this analysis (year 1 and 2 in Vadodara and Mumbai). The scores are normalized relative to the distribution of the pre-test score in the comparison group in each city and year[6]. The appendix

[6]Scores are normalized for each standard, year, and city, such that the mean and standard deviation of the comparision group is zero and one, respectively. (We subtract the mean of the control group in the pre-test, and divide by the standard deviation.) This allows for comparison across samples, as well with results from other studies.

tables 1 to 5 show the raw scores as well as the percentage of children who correctly answering the questions in the test relating to the competencies in each standard.

The randomization appears to have been successful for both programs in all years, with the exception of the Computer assisted learning program in year 3: Neither in Mumbai nor in Vadodara are there any large or systematic differences between the pre-test score and the post-test score. Except for the computer assisted learning in year 3 in Vadodara, none of the differences between the groups prior to the implementation of the program are significant. Table 5, which implement bootstrap tests of equality of distributions, confirms this pattern. The first row in table 5 present the p. value for the hypothesis that the two distribution are equal, the second row present the p. value for the hypothesis that the treatment distribution stochastically dominates the control distribution. The third line present the p. value for the hypothesis that the control distribution stochastically dominate the treatment distribution. In the pre test scores, the distributions between treatment and control can never be statistically distinguished from each other, except in the case of the CAL program in year 2.

The raw scores, and the percentage of children correctly answering the questions relating to the curriculum in each standard give an idea of how little these children actually know.[7] In standard three in Vadodara in the second year, for example, the average student in math scores about 16%, both in the control and treatment groups. Since one math question is multiple-choice, on average a student who knows nothing will score 1.8% points. If a student can consistently order two numbers and add two single digit numbers, she earns the additional 14% needed to achieve the average third standard performance. Only 5.4% of third standard children in Vadodara pass the standard 1 competencies in maths in standard 3 in Vadodara (and 14% in Mumbai). Standard one competencies cover number recognition, counting and one digit addition and subtraction.

The results are more encouraging in verbal competencies: 50% of the standard 3 children pass the standard 1 competencies in Vadodara (reading a single word, choosing the right spelling among different possible spelling for a word), and 65% do so in Mumbai.

[7] The full results for this are in the appendix tables.

## 4.2 Attrition and Transfers

Table 6A and 6B present the levels of attrition in Mumbai and Vadodara for both programs. We present attrition that occurred between the pre-test and post-test for both cities in both years, as well as the two-year attrition (in Mumbai, for standard 4 only), broken down by treatment status. To minimize attrition, the survey team visited children who were not present at the post-test in their home, and administered the test then.

Attrition was generally very low, except for Vadodara in year 1. The high attrition in this year is likely attributable to the civil unrest (severe riots affected the city in 2002). The post-test was conducted after the riots; the research team attempted to track down all of the children who did not appear for the exam, but many had left for their village during the riots. Attrition rates are not different in the comparison group than in the treatment group: In year 1 in Vadodara, attrition was 19% in the balsakhi treatment group, and 18% in the comparison group. In year 2, attrition was 4% in the balsakhi and the non balsakhi groups comparison group. In Mumbai year 1, attrition was 7% in the treatment group, and 7.5% in the comparison group, while in year 2 it was 7.7% in the treatment group and 7.3% in the comparison group. In the CAL program, attrition was 4.1% in the treatment group and 4.8% in the control group in year 1, and 7.3% and 6.8% respectively in year 2.

The fact that there was no differential attrition rate in the treatment and control groups suggests that the estimate of the treatment effects will not be biased, unless different types of people drop out from the sample in the treatment and the control groups (Angrist, 1995). This does not seem to occur in our study: 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 column of each sample group 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 most cases. In Mumbai in the second year, there is some evidence that the attritors may have had worst pre-test scores than the stayers in the treatment group, compared to the comparison group. In the CAL program in year 2 and year 3, we find the opposite, with the attritors in the treatment group seeming to perform better than the non-attritors (which is different from all the other tables). This will tend to bias the results obtained from simple differences upwards (the effect on the difference in difference estimate and the lagged dependent variables specification are unclear), although since the attrition is very low, this is unlikely to have a large effects.

Finally, both the attrition and the difference in test scores are also similar among the bottom 20 children in each school, the group of children who were the most likely to be assigned to a balsakhi (these results are not reported to save space).

In what follows, the treatment status of a child will be assigned to him as a function of the school based upon the school in which they took the pre-test. If students could transfer, this could theoretically introduce two sources of bias. First, if students were able to transfer prior to the pre-test, then treatment schools may have gained students likely to experience a significant improvement in test scores over the following year, generating a positive bias. Second, if motivated students transferred during the academic year, then some of the control group would have experienced the treatment causing us to underestimate the treatment effect.

These biases, however, do not affect our estimates. The program was not announced prior to the start of the school year. In addition, parents seem to rarely inquire about programs offered through the school. And even if they were interested, school transfers are very unlikely in both Baroda and Bombay. Administrators provide them only reluctantly, and parents have a limited number of alternative schools. Most areas have only a few schools of the same medium in the same area. Finally, since we were sensitive to the potential problems that could arise due to transfers, we also checked for students that took the pre-test in a control school and the pre-test in a treatment school and found none.

# 5 Effects of the Balsakhi and the CAL Programs

## 5.1 Attendance

Part of the goal of the Balsakhi program was to make it easier for parents to play a role in their children's education, by serving as an intermediary between parents and the school environment. One could therefore have expected an impact of the program on attendance. In practice, there does not seem to be any: table 13 shows the effect of the program of attendance in both cities (attendance was no collected in year 1 in Vadodara). In no cities and no class do we see any impact of the program on attendance.[8]

The CAL program could have affected attendance as well, by making school more attractive for students, at least on days where they are scheduled to go to the CAL program. Table 11 also shows the effect of the CAL program on attendance. In year 2 (the first year of the program), CAL appears to have no effect on attendance. In year 2, we see a small, positive effect, significant at the 10 percent level (the effect 2.5 percentage points, with a standard errors of 1.5 percent). This may indicate a small effect of the CAL program on attendance, but the confidence in this result may be limited by the fact that the treatment schools appeared to be doing better in term of test scores before the program.

### 5.2 Test scores: Balsakhi program

Tables 2 and 3 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.

The Balsakhi program appear to be succesful: In all years and standards, for both tests, and in both cities, and for all subgroups, the difference in post-test scores between treatment and control groups is positive. The hypothesis that the two distribution are equal can be rejected in both cities and grade in year 2 at 95% level of confidence, and at the 11% level of confidence in standard 3 in Vadodara (in standard 4 in Vadodara and standard 3 in Mumbai the hypothesis cannot be rejected). The hypothesis that the control group distribution stochastically dominates the distribution of the treatment group can always be rejected at least at the 10% level, while the hypothesis that the distribution in the treatment group dominates that in the control group can never be rejected.

In the first year in Vadodara (table 2), the difference in post-test score between treatment and control groups was 0.18 standard deviations in standard three for math, .16 in standard 3 for language, and .16 and .09 in standard 4, for math and language respectively. Note that between the mid test and the post test, scores have actually declined in year 1: this is likely due to the riots, which severely disturbed the schools and the children[9] The results in Mumbai

[8]The data on attendance is obtained by roster calls at un-announced visits. In Mumbai, we also collected attendance from roster filled by the teachers. They generally show a higher attendance rate, and there appear to be no difference between attendance in treatment and comparison school using this measure.

[9]Throughout the paper, test results and program effects are presented in terms of standard deviations, unless

(table 4) are remarkably similar, with the math and language test scores improving by 0.16 and 0.15 standard deviation, respectively.

In the second year of the program, the effects are larger: In Vadodara (table 2, the difference in total test scores is .44 for math and 0.25 for language in standard three, and .34 and 0.30 in standard four, for math and language respectively. In Mumbai in year two (table 3), the IV estimate of the impact of the program on test scores differences in are .26 and .11 in standard 3 (for math and language respectively) an .49 and .20 in standard 4 (for math and language respectively). In year two in Vadodara, all of the differences between treatment and control groups are statistically significant, while for Mumbai, the standard four results are significant.

Because test scores have a strong persistent component, the precision of these estimates can be improved significantly, however, by controlling for the child's pre-test score (equation 5 or turning to a differences-in-differences specification (equation 6). Since the randomization appeared to be successful, and attrition was low in both the treatment and comparison groups, the point estimates should be similar in the simple differences and these two specifications. Table 7 presents the results, in various years, cities, standards, and sub-groups. For Mumbai in year 2, we estimate the treatment effect in two ways: first, we instrument for the dummy indicating whether or not the school received a balsakhi with a dummy for whether the school was assigned to the treatment group; second, we include only schools that got a balsakhi in at least one standard in the sample. The estimates using either specification are very similar.

As expected, the point estimates suggest a substantial treatment effect, and the standard errors are lower than the simple differences. Pulling both cities and standard together (in the first two rows of table 7), the impact of the program was 0.14 standard deviation overall in the first year, and 0.27 standard deviation in the second year (0.28 using the value added specification). All estimates for total score are significant at the 99% confidence level.

The impact is bigger in the second year, and bigger for math than for language in both years (0.19 standard deviations versus 0.069 in the first year, and 0.32 versus 0.15 standard deviations in the second year; all but first-year verbal scores are significant at the 99% level.) For both years and both subjects pooled, the effect are a little larger in Vadodara than in Mumbai (with a total-score effect of 0.14 standard deviation versus 0.12 in the first year (standard 3 only),

otherwise specified.

and 0.31 versus 0.20 in the second year (both standards)). The difference is the strongest for language, where there is a significant impact in both years for Vadodara (0.11 and 0.23 standard deviation respectively), but no significant impact in either year in Mumbai for grade 3 (0.06 standard deviation for standard 3 in year 1, and 0.051 standard deviation in year 2), though the effect is larger and significant for grade 4 (0.14 standard deviation). For both cities and both subjects, the effects are very similar in standard 3 and standard 4. We also computed all those estimates for both genders separately, and found the impact to be very similar (results not reported).

In the last panel of the table, we display our estimate of the impact of the program for two years in Mumbai (for children who were in a treatment school in standard 3, and stayed in the treatment school).[10] First, it appears that the effect of the first year does not seem to persist over the summer: at the pre-test in year 2, children who were in a treatment class in year 1 do not seem to know more than children who weren't. However, the effect of two years of treatment (from year1 pre-test score to year 2 post-test score) is substantially larger than that of either individual year (0.60 standard deviation in math, for example, versus 0.40 for year 2 in grade 4): it seems likely that the foundation laid in the first year of the program helped the children benefit from the second year of the program.

Compared to the other educational interventions, this program thus appear to be quite effective. The Tennessee STAR experiment, for example, where class size was reduced by 7 to 8 children (from 22 to about 15), improved test scores by about 0.21 standard deviation. This program improved test scores by 0.27 standard deviations in the second year, by reducing effective class size from 40 to 20 children on average (averaging over the balsakhi and the non-balsakhi group) for part of the day, but doing so by hiring an assistant paid a fraction of the teacher's salary.

### 5.3 Test scores: Computer Assisted Learning

Table 4 shows the simple difference in the mid and post test in the CAL program. The math test scores are significantly greater in the treatment schools in the post test in both years. In

[10]It was not possible to do that in Vadodara, because the riots at the end of year 1 led to a massive churning of students in the school.

year 2, maths post-test score is on average 0.33 standard deviation higher in the CAL schools (with a standard deviation of 0.087). In year 3, it is 0.63 standard deviation higher, but we need to take into account the fact that the pre test scores were already 0.15 higher in year 3.

This is done explicitly in table 8, which shows the difference in difference and value added specification of the effect of the CAL program. The results are confirmed: The CAL program has a strong effect on math score (0.36 standard deviations in the first year, and 0.51 standard deviation in the second year, using the value added specification). It has no discernable impact on language scores (the effects are very close to being exactly zero). This is not surprising, since the software targeted exclusively math skills., although one could have expected some spillover effects on language skills (through increased attendance or the practice of reading instructions). The effect on the sum of language and math test scores is 0.18 standard deviation in year 2, and 0.19 standard deviation in year 3.

Panel B explicitly compares the Balsakhi and the CAL effects, and examine their interactions, in the year where they were implemented at the same time (the randomization was stratified). When not interacted, CAL has a larger effect on math test scores than Balsakhi (although the difference is not significant) and a smaller effect on overall test score (although the difference is not significant either). The programs appear to have no interaction with each others: the coefficients on the interaction on the maths and overall test score are small, insignificant, and negative.

### 5.4 Effect on specific competencies and distributional effects

The balsakhi program was primarily intended to help children at the lower end of the ability distribution, by providing especially targeted instruction. However, as we already mentioned, it could still help the higher scoring children, either because they are assigned to the balsakhi, or because they benefit from smaller classes when their classmates are with the balsakhi.

At worst, the program could have increased the average score while hurting children at the bottom of the distribution (by reducing class size by sending the most disturbing children away).

However, it does not appear to be the case: table 5 shows that the distribution of the test score in the treatment schools stochastically dominate the distribution of the test score in the comparison schools in all projects where the simple average was significant. Figures 1 show one example of how the distributions transformed, for Vadodara year 2. While the distributions are not distinguishable in year 1, they are very different in year 2, with the distribution in the treatment schools clearly dominating the distribution in year 2.

Table 9 offers more details on the level at which both programs was effective.[11] Estimates in this table suggest that, for math, the biggest effect of the balsakhi program was on the competencies of standard 1: in Vadodara for example, the program increased the fraction of children who mastered the competencies of the first standard in math by 4.0% in the first year, and 7.3% in the second year. In Mumbai the effect was 4.5% and 13.1% respectively. The effect on the fraction of children demonstrating knowledge of standard 3 competencies is much smaller. In language, the most important effect seems to be to help children master the competencies of standard 2 This may not be surprising, since many children seemed to have already mastered the competencies of standard 1. The effect of the program may thus be the strongest on the easiest competencies not already mastered by many pupils. These results correspond well with the stated role of the program, which was to work with children on basic competencies.

The CAL program affected only math competencies, and seem to have had an equal effect on the number of children able to pass standard 1 and standard 2 competencies (about 13 percentage points for each in year 1). It also affected standard 3 competencies, especially in year three (it increased the fraction of students that passed them by 7.9 percentage points, when the fraction of fourth-standard student who passed these competencies in the pre-test was only 1.3%). The CAL program, unlike the balsakhi program, thus appear to have the potential to help children at all levels.

In figure 3 present estimate of the post-test scores as a function of the pre-test score rank in the overall distribution (using a Fan locally weighted regression) for treatment and control schools in year 2 (both cities and grades are pooled). Children do on average better on the post test in treatment schools than in comparison schools for any level of the pre-test. In figure 4, we present an estimate of the treatment effect (using the value added specification) at different centile of the initial test score distribution, for the balsakhi program and the CAL program (we present both effects for year 2 and standard 4, for consistency). We also present the probability

[11] To save space, these estimates are presented only for the lagged dependent variable specification. The difference in differences specification delivers very similar results.

to be sent to the balsakhi for a child at this centile in the percentile distribution. The estimates of the treatment effect are obtained as the difference between the functions presented in figure 3. The estimate of the probability to be assigned to the balsakhi are estimated with a locally weighted regression. The effect of the Balsakhi program is smaller for children who were doing better initially. This corresponds to the objective of both programs: the Balsakhi program was mean to help the children who were lagging behind, while the CAL program was meant to adapt to the child and help them progress whatever their initial level was.

The effect of the Balsakhi program at various level of the pre-test score distribution seems to follow closely the probability to be assigned to the Balsakhi. These results lead to our next question: to what extent is the program effect due to a direct effect of the balsakhi teacher (affecting only the children who got assigned to the balsakhi group) and to an indirect effect, affecting children who were *not* assigned to the balsakhi group. The fact that both the program impact and the probability of being assigned to a balsakhi declines with a child's position in the test score distribution suggest that the impact of the program may have been larger for those who were actually assigned to the balsakhi (otherwise, one would see a positive treatment effect even for children with very low probability to be assigned). However, an alternative explanation for this pattern is that the direct (or indirect) effects of the program are lower for children with higher pre-test scores, in ways that exactly tracks the decrease in the probability to be assigned. This question is further investigated in the next section.

## 6 Inside the box: direct and indirect effects

Estimating equations 3 and 6 generates estimates of the average impact of the program on all children who whose standard-school received a balsakhi. The program may impact the children in a treated school in two ways: directly, for children who were assigned to work with the balsakhi, or indirectly, because the weakest children are removed from the classroom for part of the day. This indirect effect can potentially work through two mechanisms: through a reduced number of students in the class (class size effect), and through the higher average quality of their classmates (tracking effect).

## 6.1 Statistical Framework

To separate the direct remedial education effects and the indirect effects, an ideal experiment would have identified the children who would work with the balsakhi in all schools, *before* randomly assigning treatment and comparison groups (and to not allow substitution after the initial allocation). The balsakhi effect could then be estimated by comparing children at risk of working with the balsakhi in the treatment and the comparison group. The indirect effect would have been estimated by comparing the children who were not at risk of working with the balsakhi in the treatment and the comparison group. Unfortunately, this design was not practical in this setting.

We do know, however, that the assignment to the balsakhi group was based in part on pre-test score, and that a maximum of twenty children per school in Vadodara, and twenty per class in Mumbai were assigned to a balsakhi. We use this fact to implement two different empirical strategies to disentangle direct and indirect effects.

### 6.1.1 Exploiting Assignment Probabilities

This strategy is directly inspired by figure 4, which suggests that the effect of the program closely track the probability to be assigned to the Balsakhi, which suggest that the entire effect of the program goes through the assignment.

We start by estimating assignment probability flexibly in the treatment schools as a function of the rank in the pre-test score distribution.

$$P_{ijg} = \pi_1 + \pi_2 Q_{ij} + \pi_2 Q_{ij}^2 + \pi_3 Q_{ij}^3 + \pi_4 Q_{ij}^4 + \omega_{ijg} \tag{9}$$

Denote $M_{ij}$ the vector $[1 Q_{ij} Q_{ij}^2 Q_{ij}^3 Q_{ij}^]$.

We then estimate the treatment effect as a function of the same variables, interacted with the treatment status of the schools.

$$y_{ijgPOST} = \theta y_{ijgPRE} + M_{ij}'\lambda + D_{ij} * M_{ij}'\mu + \tag{10}$$

Equation 9 and 10 form the first stage and the reduced form, respectively, of the following structural form equation:

$$y_{ijgPOST} = \theta y_{ijgPRE} + \gamma D_{ijg} + \delta P_{ijg} + M'_{ij}\alpha + \epsilon_{ijg} \quad (11)$$

The four instruments allows to estimate $\gamma$ and $\delta$. Under the maintained assumption that the indirect treatment effect $\gamma$ is constant, an overidentification test allows us to test whether the remedial education treatment effect is indeed constant.

### 6.1.2 Exploiting the non-linearity in assignment rules

The strategy in the previous subsection relies on assumptions about the relationship between the relationship between the direct and indirect effects and the initial test scores. To complement it, we implement a strategy which does not rely on this assumption.

The strategy exploits the discrete change in assignment probability for a children of rank 20 in a given class. It estimates direct remedial education and indirect class size or tracking effect for children whose test scores could place them either below rank 20 or above rank 20, depending on their class size. Estimating this parameters does not require to make any assumption about the constancy or the regularity of the direct and indirect effect at rank 20.

In schools in the treatment group, we start by predicting assignment to the balsakhi as a function of the number of students (in the school in Vadodara, in the class in Mumbai), the sum of the math and verbal score at the pre-test, and a variable indicating whether the child is among the bottom 20 children in his group.

$$P_{ijg} = \pi_1 + \pi_2 S_{ijg} + \pi_3 y_{ijgPRE} + \pi_4 R_{ijg} + \pi_5 Z_{ijg} + \omega_{ijg} \quad (12)$$

where $S_{ijg}$ is the number of student in the class or the school, $y_{ijgPRE}$ is the score of the child at the pre-test, $R_{ijg}$ is the rank of the child in the class (starting from the bottom), and $Z_{ijg}$ is a dummy indicating whether the child is among the bottom 20 children in the class. We will show that, even after controlling linearly for the class rank, the dummy $Z_{ijg}$ predicts whether or not the child was assigned to the balsakhi.

Denoting $X_{ijg}$ the vector $[S_{ijg} \quad y_{ijgPRE} \quad R_{ijg}]$, the following equation (which interacts the variables in equation 12 with a dummy for whether the child is in the balsakhi group) predicts assignment to the balsakhi in the whole sample.

$$P_{ijg} = \alpha + \gamma D_{ijg} + \beta \left(Z_{ijg} * D_{ijg}\right) + \mu Z_{ijg} + X'_{ijg}\kappa + \lambda \left(X'_{ijg} * D_{ijg}\right) + \epsilon_{ijg} \tag{13}$$

We can then regress the post test scores on the same variables (controlling for pre-test score), and examine whether being one of the bottom 20 children is associated with a bigger effect for those whose school is in the treatment group:

$$y_{ijgPOST} = \alpha + \beta Z_{ijg} * D_{ijg} + \gamma D_{ijg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda\epsilon ijg \tag{14}$$

Equation 14 and 13 form the first stage and the reduced form of a instrumental variables estimation of the following equation:

$$y_{ijgPOST} = \alpha + \beta P_{ijg} + \gamma B_{ijg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda + \epsilon_{ijg} \tag{15}$$

where $D_{ijg}$ and $Z_{ijg} * D_{ijg}$ are the excluded instruments. The identification assumption underlying this estimation strategy is that the only reason why the treatment effect varies with the variable $Z_{ijg}$ is because $Z_{ijg}$ makes it more likely that the child is sent to the balsakhi group. However, the effect of the treatment is allowed to vary with class size, the test score, and the rank of the child. We also estimate an alternative specification which controls for a fourth-order polynomial in the rank of the child. In this equation, the effect of being assigned to the balsakhi group is given by $\beta + \gamma$, and the effect of being in a balsakhi school, but not assigned to the balsakhi group, is given by $\gamma$.

### 6.2 Results

As we explained above, we propose using a dummy for whether a child belongs to the bottom 20 children of a class as an instrument for whether he is assigned to the balsakhi group. Columns 1 to 3 in table 10 show that in both Mumbai and in Vadodara, a dummy for whether a child belongs to the bottom 20 in his class predicts his assignment to the balsakhi, even after controlling for her rank, her score at the pre-test, and the number of students in her class (which are all negatively and significantly associated with assignment to the balsakhi). Not surprisingly, because some schools in Bombay were not assigned a balsakhi, all coefficients are smaller. In columns 4 to

6, we present the reduced form estimates for test score gain. The coefficient of the interaction between the dummy for belonging to the bottom 20 children in the class and belonging to a treatment school is significant in all of these columns, which indicates that, conditional on being in a school assigned to the treatment group, the treatment effect is actually bigger if the child is more likely to be assigned to the balsakhi.

In table 11, we present instrumental variables estimates of the direct and indirect impact of being in a balsakhi group, using the two strategies described above. The first 3 columns use the treatment dummy (which is 1 for every child assigned to a treatment-grade with a balsakhi) and its interactions with the pre-test score, and its square, cubic and quartic as an instruments for the Balsakhi school and balsakhi assignment variables. The last lines in the table show the F stat for the excluded interactions, which are highly significant, and the p value for the overidentification test.

The estimates strongly reject the hypothesis that being in a balsakhi school has any effect for children who were not themselves sent to the balsakhi: The effect of the program seem to be concentrated on children who were indeed assigned. The effect on these children is large: they gain 0.6 standard deviation in the overall test scores (which is over half of what was gained by the children in one year for year 2). We cannot reject the hypotheis that the treatment effect is constant: the fact that the balsakhi program affects mostly children at the bottom of the test score distributions simply reflects the fact that the children at the bottom of the test score distribution are more likely to be affected to the balsakhi group. It is important to keep in mind that this estimate is local: it is likely that none of the children assigned to the balsakhi program were very advanced, and the balsakhi program is not adapted to teach advanced competencies.

Columns 4 to 6 present the estimate of the program effect using the discontinuity in the assignment rule at rank 20. These estimates also reject the hypothesis that the program had any effect on children who were not send to the balsakhi. The point estimate of the direct effect (the effect to work with a balsakhi) are even larger than previously (1 standard deviation), but they are also less precise and cannot be statistically distinguished from the estimate in column 1 to 3.

Both strategies lead to the same conclusions: the direct effect of the balsakhi program is very large, but the reduction in class size induced by the program had no direct effect. Since the average class size is 45 (though the average student has 63 students in his class), the reduction in class size was about the same in the balsakhi group and the non-balskahi group, and the effect can therefore be directly compared. They suggest that the same reduction in class size would be much more effective if it was done through a balsakhi type program than if twice as many additional teachers were hired.

Table 12 investigates whether the direct and indirect impacts of the balsakhi program vary with school characteristics.

The first characteristics we consider is class size. While the size of the balsakhi group is always 20, the size of the remaining group is linked to the original balsakhi group. The reduction in the class size in the balsakhi group is thus larger in big school, while the proportional reduction in the size of the class for the children who do not attend the balsakhi program is smaller in big schools. We divide the schools into two groups, those where with more or less than 40 children. In small schools, the balsakhi group is actually larger than the size of the group that remains in the regular classroom. As expected, the effect of the balsakhi program appear to be about twice as large in large schools than in small schools, though the estimates are noisy. The effect of the program on unassigned children is smaller in large schools. The interaction is negative and, and it is significant when we use the more precise strategy. They also suggest that children in small schools may have benefited from the class size reduction (the effect is 0.2 standard deviation, and is significant in the combined sample). This suggest that class size reduction may be effective if they result in very small class size (less than 20) for the regular teacher. This may make it possible for teachers to change the way they normally teach.

The two other characteristics we consider, variance in initial test score and average test score of the bottom 20 children, attempt to capture the possibility of a benefit to tracking. However, neither the variance in initial test score nor the average test score of the bottom 20 children appear to have an impact on either the direct or indirect effects.

## 7 Cost Benefit Analysis

In seeking to improve the academic performance of schoolchildren, governments could potentially hire additional teachers, hire balsakhis or put computers in classrooms.

Since we do not detect any effect of reducing class size on test scores, hiring new teachers (who are paid several times more what balsakhis are paid) does not appear to be a cost-effective strategy. Even using the most optimistic estimate that hiring teachers may have increased scores in small schools by 0.2 standard deviation, hiring balsakhis would be several times more cost effective than hiring teachers.

We can also compare the cost of one year of the balsakhi program with one year of the CAL program. The cost per student per year of the balsakhi program is 107 rupees, or approximately 2.25 dollars. The recurring expenditures of the CAL program are 367 rupees, but the cost of the CAL program including the start-up costs of the computers and software (assuming they are depreciated over five years) is 722 rupees. Thus, using the estimates from table , we can calculate the relative cost effectiveness of each program. CAL increases the math score by 0.41 standard deviations and the overall test score by 0.25 whereas the Balsakhi program increases the math score by 0.31 and the total score by 0.28. Since CAL costs 6.7 times as much as the Balsakhi program per student, the Balsakhi program is 5 times more cost effective for math and 7.7 times more cost effective for the total score. However, the cost benefit analysis needs to keep in mind that the computer assisted learning program may be more effective at teaching more advanced competencies.

## 8 Conclusion

This paper reports the results of a remedial education and a computer assisted learning programs.

The 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 whose schools benefited from the program improved by 0.14 standard deviations in the first year, and 0.28 in the second year. We also estimate that children who were directly affected by the program improved their test scores by at least 0.6 standard deviation in the second year.

A computer assisted learning program provided each child in the fourth standard with two hours of shared computer time per week, in which students played educational games that reinforced mathematics skills. Contrary to what was found in developed countries settings, the program was also very effective, increasing math scores by 0.36 standard deviation the first year, and 0.54 the second year.

These results show that it is possible to dramatically increase the quality of education in urban India, a very important result since less than a third of Indian school children can read when they leave school. However, this is not likely to be achieved by simply increasing resources without changing the way teaching is conducted.

## References


Angrist, Joshua D. (1995) “Conditioning on the Probability of Selection to Control Selection Bias,” NBER Technical Paper #181

Angrist, Joshua, et. al. ”Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment.” American Economic Review 92 (2002): 1535-1558

Banerjee, Abhijit, Suraj Jacob and Michael Kremer. ”Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials.” MIT Department of Economics Working Paper (2002): .

Chin, Amiee (2002) ”The Returns to School Quality When School Quality is Very Low: Evidence from Operation Blackboard in India.” University of Houston

Duflo, Esther, (2001). ”Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment.” American Economic Review 91 (2001): 795-813

Glewwe, Paul, Michael Kremer and Sylvie Moulin (1997) ”Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya.” Manuscript, Harvard University Department of Economics

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz. (2003) ”Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya.” Journal of Development Economics, forthcoming

Kremer, Michael and Ted Miguel (2003) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica, forthcoming

Kremer, Miguel, and Thornton (2002) "Merit Scholarships," manuscript, Harvard Economics.

Spohr, Chirstopher. (2003) "Formal schooling and workforce participation in a rapidly developing economy: evidence from 'compulsory' junior high school in Taiwan." Journal of Development Economics 70: 291-321

Vermeersch, Christel (2002) "School Meals and Educational Achievement: Evidence from a Randomized Evaluation,." Manuscript, Harvard University Department of Economics