REMEDYING EDUCATION: EVIDENCE FROM TWO RANDOMIZED EXPERIMENTS IN INDIA*

Abhijit V. Banerjee
Shawn Cole
Esther Duflo
Leigh Linden

This paper presents the results of two randomized experiments conducted in schools in urban India. A remedial education program hired young women to teach students lagging behind in basic literacy and numeracy skills. It increased average test scores of all children in treatment schools by 0.28 standard deviation, mostly due to large gains experienced by children at the bottom of the test-score distribution. A computer-assisted learning program focusing on math increased math scores by 0.47 standard deviation. One year after the programs were over, initial gains remained significant for targeted children, but they faded to about 0.10 standard deviation.

I. Introduction

The recent World Development Report on "Making Services Work for Poor People" [World Bank 2004] illustrates well the essential tension in the public conversation about primary education in developing countries. On the one hand, the report embraces the broad agreement, now enshrined in the Millennium Development Goals, that primary education should be universal. On the other hand, it describes in detail the dismal quality of the educational services that developing countries offer to the poor.

For example, a 2005 India-wide survey on educational attainment found that 44 percent of the children aged 7–12 cannot read a basic paragraph, and 50 percent cannot do simple subtraction [Pratham 2005] even though most are enrolled in school. Even in urban India, where widespread absenteeism by students and

* This project was 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 requests for new data: Pratima Bandekar, Rukmini Banerji, Lekha Bhatt, Madhav Chavan, Shekhar Hardikar, Rajashree Kabare, Aditya Natraj, and many others. We thank Jim Berry, Marc Shotland, Mukesh Prajapati, and Nandit Bhatt for their excellent work coordinating the fieldwork and for their remarkable work in developing and improving the CAL program. Kartini Shastry provided superb research assistance. Two editors and three referees provided very useful comments. We also thank Joshua Angrist, Angus Deaton, Rachel Glennerster, Michael Kremer, Alan Krueger, Victor Lavy, and Caroline Minter-Hoxby for their comments. For financial support, we thank the ICICI corporation, the World Bank, the Alfred P. Sloan Foundation, and the John D. and Catherine T. MacArthur Foundation.

The Quarterly Journal of Economics, August 2007

teachers is not an issue, the learning levels are very low: in Vadodara, a major Indian city and a site for the study in this paper, only 19.5 percent of the students enrolled in grade 3 can correctly answer questions testing grade 1 math competencies.

In these conditions, policies that promote school enrollment may not promote learning. And indeed, the recent evidence suggests that many interventions, which increase school participation, do not improve test scores for the average student.[1] Students often seem not to learn anything in the additional days that they spend at school.[2]

It is therefore clear that efforts to get children into school must be accompanied by significant improvements in the quality of schools that serve these children. The problem is that while we now know a reasonable amount about how to get children into school, much less is known about how to improve school quality in a cost-effective way. Worse still, a number of rigorous, randomized evaluations have confirmed that spending more on resources like textbooks [Glewwe, Kremer, and Moulin 2002], flip charts [Glewwe et al. 2004], or additional teachers [Banerjee, Jacob, and Kremer 2004] has no impact on children's test scores (see Glewwe and Kremer [forthcoming] for discussions and more references). These results have led to a general skepticism about the ability of interventions focusing on inputs to make a difference (echoing Hanushek's [1986 and 1995] earlier assessment for both the U. S. and developing countries) and have led many, including the above-mentioned World Development Report, to advocate more systemic reforms designed to change the incentives faced by teachers, parents, and children.

It is not clear, however, that we know enough to entirely give up on inputs. Based on existing evidence, it remains possible that additional inputs actually can work but only if they address specific unmet needs in the school.

Ironically, the difficulty in improving the quality of education may in part be a by-product of the success in getting more children to attend school. Neither the pedagogy nor the curriculum has been adapted to take into account the influx of children and their characteristics: many of these children are first generation

1. These include giving children deworming drugs [Miguel and Kremer 2004] and providing school meals for children [Vermeersch and Kremer 2005].

2. This is true when evaluating only children who were enrolled before the intervention, suggesting this result is not due to a change in the composition of the children.

learners whose parents are not in a position to follow what is happening in school or to react if their child falls behind. Yet, in many countries, the school system continues to operate as if it were catering to the elite. This may explain why just providing more inputs to the existing system or more school days is often ineffective. For many children, neither more inputs nor an extra day makes much of a difference because what is being taught in class is too hard for them. For example, Glewwe, Kremer, and Moulin [2002] found that new textbooks make no difference for the test scores of the average child but do help those who had already done well on the pretest. The authors suggest that this is because the textbooks were written in English (the language of instruction, in theory), which for most children is the third language.

Taken together, these results suggest that inputs specifically targeted to helping weaker students learn may be effective.

This paper reports the results from randomized evaluations of two programs that provide supplementary inputs to children in schools that cater to children from poor families in urban India. The first intervention is specifically targeted to the weakest children: it is a remedial education program, where a young woman ("Balsakhi") from the community works on basic skills with children who have reached grade 3 or 4 without having mastered them. These children are taken out of the regular classroom to work with this young woman for two hours per day (the school day is about four hours). The second intervention is addressed to all children but is adapted to each child's current level of achievement. It is a computer-assisted learning program where children in grade 4 are offered two hours of shared computer time per week during which they play games that involve solving math problems whose level of difficulty responds to their ability to solve them. Both programs were implemented by Pratham, a very large NGO operating in conjunction with government schools in India. The remedial education was run in Mumbai (formerly known as Bombay) and Vadodara (formerly known as Baroda), two of the most important cities in western India. The Computer-Assisted Learning Program was run only in Vadodara.

In contrast to the disappointing results of the earlier literature, we find that both programs had a substantial positive effect on children's academic achievement, at least in the short run. This is true in both years and cities, despite the instability of the environment (notably major communal riots in Vadodara in 2002, which severely disturbed the schools).[3] The remedial education program increased average test scores in the treatment schools by 0.14 standard deviations in the first year, and 0.28 in the second year. Moreover, the weaker students, who are the primary target of the program, gained the most. In the second year, children in the bottom third of the initial distribution gained over 0.40 standard deviations. Using an instrumental variable strategy, we estimate that the entire effect of the remedial education program derives from a very large (0.6 standard deviations) improvement of the children within the classroom who were sent for remedial education. In contrast, there is no discernible impact on their classroom peers, who were "treated" with smaller class sizes and a more homogenous classroom, consistent with the previous literature suggesting that inputs alone are ineffective.

The computer-assisted learning increased math scores by 0.35 standard deviations the first year, 0.47 the second year, and was equally effective for all students.

Such large gains are short-lived, although some effect persists over time: One year after leaving the program, initially low scoring students who were in balsakhi schools scored approximately 0.1 standard deviations higher than their control-group peers. Students at all levels of aptitude performed better in math (0.1 standard deviations) if they were in schools where the computer-assisted math learning program was implemented.

The remainder of the paper is organized as follows. In Section II, we describe the remedial education and computer-assisted learning interventions in detail. Section III describes the evaluation design. In Sections IV and V, we present the short- and longer-run results (respectively) of the evaluation. In Section VI, we attempt to distinguish the effect on those who were taught by a remedial education instructor from the indirect effect on those who remained with the original instructor, hence enjoying a smaller and more homogenous classroom. Section VII concludes.

3. A train carrying Hindus traveling to a controversial site (where a mosque had been destroyed by a Hindu mob in 1991) caught fire in February 2002, allegedly because of an attack by Muslims. Many Muslim communities were attacked in retaliation during the next several weeks in major cities in Gujarat, causing hundreds of casualties and major disorder.

II. The Programs

The interventions evaluated in this study were implemented in conjunction with the Indian organization Pratham. Pratham was established in Mumbai in 1994 with initial support from UNICEF and has since expanded to several other cities in India. Pratham now reaches over 200,000 children in fourteen states in India, employing thousands. It works closely with the government: most of its programs are conducted in the municipal schools or in close collaboration with them, and Pratham also provides technical assistance to the government.

II.A. Remedial Education: The Balsakhi Program

One of Pratham's core programs at the time of this study was a remedial education program, called the Balsakhi Program (balsakhi means "the child's friend"). This program, in place in many municipal schools, provides government schools with a teacher (a "balsakhi," usually a young woman, recruited from the local community, who has herself finished secondary school) to work with children in the third and fourth grades who have been identified as falling behind their peers. While the exact details vary according to local conditions, the instructor typically meets with a group of approximately 15–20 children in a class for two hours a day during school hours (the school day is about four hours long). Instruction focuses on the core competencies the children should have learned in the first and second grades, 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 by Pratham in many Indian cities, reaching tens of thousands of students, and by Pratham in collaboration with state governments, reaching hundreds of thousands. It was started in Mumbai in 1998 and expanded to Vadodara in 1999.

An important characteristic of this program is the ease with which it can be scaled up. Because Pratham relies on local personnel, trained for a short period of time, the program is of very low cost (each teacher is paid 500–750 rupees, or 10–15 dollars, per month) and is easily replicated. Indeed, though we evaluated the program in only one subdivision of Mumbai ("L Ward"), the intervention was programmatically identical to Pratham's interventions in many other wards of Mumbai. The curriculum and the pedagogy are simple and standardized. There is rapid turnover among the balsakhis (each stays 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.

These characteristics distinguish the program from standard remedial education programs in the developed world, which tend to use highly qualified individuals to provide small-group or individual instruction.[4]

II.B. Computer-Assisted Learning

The Computer-Assisted Learning (CAL) Program takes advantage of a policy put in place by the government of Gujarat. In 2000, the government delivered four computers to each of the 100 municipal government-run primary schools in the city of Vadodara (80 percent of the schools).

The idea of using computers to remedy the shortage of qualified teachers is very popular in Indian policy circles. Computers have the potential to both directly improve learning and indirectly increase attendance by making school more attractive. Unfortunately, there exists very little rigorous evidence on the impact of computers on educational outcomes and no reliable evidence for India or other developing countries. The evidence available from developed countries is not encouraging: Angrist and Lavy [2002], Krueger and Rouse [2004], Machin, McNally, and Silva [2006], and Leuven et al. [2004] all find little or no effect of computerized instruction on test scores. It is not clear, however, that these results apply in developing countries, where computers may replace teachers with much less motivation and training.

In Vadodara, a survey conducted by Pratham in June 2002 suggested that very few of these computers were actually used by children in elementary grade levels. Pratham hired a team of instructors from the local community and provided them with five

4. See Lavy and Schlosser [2005] and Machin, Meghir, and McNally [2004] for two evaluations of remedial education programs in Israel and the UK, respectively. They both find small, positive effects.

days of computer training. These instructors provided children with two hours of shared computer time per week (two children shared 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, which emphasized basic competencies in the official mathematics curriculum. In the first year of the program, Pratham relied on internally developed and off-the-shelf software, and in the second year, they partnered with Media-Pro, a local software company, to develop additional software to more closely follow the Vadodara curriculum.

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 at which the CAL Program was not implemented were free to use the computers on their own, but in practice, we never found them being used for instructional purpose.

III. Evaluation Design

III.A. Sample: Vadodara

Balsakhi. The experiment began in the 2001–2002 school year (year 1), after a pilot in the previous year. To ensure a balanced sample, assignment was stratified by language, pretest score, and gender. Ninety-eight of Vadodara's 122 government primary schools participated in year 1 of the study. Half the schools (Group A) were given a balsakhi to work with children in grade 3; the other half (Group B) were given balsakhis to work in grade 4. Table I describes the design and reports the sample size of the study.

The program continued during the school year 2002–2003 (year 2). Schools in Group A, where the balsakhi was assigned in grade three in the year 2001–2002, were now assigned a balsakhi in grade 4. Schools in Group B, where the balsakhi was assigned to grade 4 in year 1, received balsakhi assistance for grade 3 in year 2. In addition, in year 2, the remaining twenty-four primary schools not previously included in the study were added by randomly assigning them Group A or B.

TABLE I
SAMPLE DESIGN AND TIME LINE

	Year 1 (2001–2002)		Year 2 (2002–2003)		Year 3 (2003–2004)	
	Grade 3	Grade 4	Grade 3	Grade 4	Grade 3	Grade 4
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Vadodara						
Balsakhi						
Group A (5,264 students in 49 schools in year 1; 6,071 students in 61 schools in year 2)	Balsakhi	No balsakhi	No balsakhi	Balsakhi	No balsakhi	No balsakhi
Group B (4,934 students in 49 schools in year 1; 6,344 students in 61 schools in year 2)	No balsakhi	Balsakhi	Balsakhi	No balsakhi	No balsakhi	No balsakhi
Computer-Assisted Learning (CAL)						
Group A1B1 (2,850 students in 55 schools in year 2; 2,814 students in 55 schools in year 3)	No CAL	No CAL	No CAL	CAL	No CAL	No CAL
Group A2B2 (3,095 students in 56 schools in year 2; 3,131 students in 56 schools in year 3)	No CAL	No CAL	No CAL	No Cal	No CAL	CAL
Panel B: Mumbai						
Balsakhi						
Group C (2,592 students in 32 schools in year 1; 5,755 students in 38 schools in year 2)	Balsakhi	No balsakhi	No balsakhi	Balsakhi	No balsakhi	No balsakhi
Group D (2,182 students in 35 schools year 1; 4,990 students in 39 schools in year 2)	No balsakhi	No balsakhi	Balsakhi	No balsakhi	No balsakhi	No balsakhi

Notes: This table displays the assignment to schools in various treatment groups in the three years of the evaluation.

Group A1B1 and A2B2 were constituted by randomly assigning half the schools in Group A and half the schools in Group B to the Group A1B1 and the remaining schools to the Group A2B2.

Schools assigned to Group A (resp. B) in 2001–2002 remained in Group A (resp. B) in 2002–2003. Twelve new schools were brought in the study and assigned randomly to Groups A and B.

Schools assigned to Group C (resp. D) in 2001–2002 remained in Group C (resp. D) in 2002–2003. Ten new schools were brought in the study and assigned randomly to Groups C and D.

Given this design, in each year, children in grade 3 in schools that received the program for grade 4 form the comparison group for children that receive the program for grade 3, and vice versa. While the assignment strategy ensures treatment and comparison groups are comparable, the estimates of the program effect would be biased downwards if the schools reassigned resources from one grade to the other in response to the program. In practice, the way schools are organized in urban India (and, in particular, in Vadodara and Mumbai) makes this extremely unlikely: schools have a fixed number of classes (a group of students and a teacher) per grade. All students are automatically promoted so that the principals have no discretion in the number of students per class or the number of teachers per grade. Most schools have just enough classrooms for each class, and in Vadodara the balsakhi class typically met outside or in a hallway. Teachers were assigned to classes before the program was implemented, and we observed no instance of subsequent reassignment to a different standard. There are essentially no other resources to speak of that the head teacher could allocate to the grade that did not receive the balsakhi. Thus, we are confident that there was no reallocation of resources to the grade that did not receive the balsakhi, which makes these students a good comparison group.

Note that this design allows us to estimate both one-year and two-year effects of the program, since a child entering grade 3 in a school where the program was offered in grade 3 in year 1 (Group A school) would remain in the treatment group in the second year, when in grade 4.

Computer-Assisted Learning. The CAL Program was first implemented in almost half of the municipal primary schools in Vadodara in 2002–2003, focusing exclusively on children in grade 4. In a few schools, computers could not physically be installed either because of space constraints or lack of electricity to run the computers. These schools were excluded from the randomization. Among remaining schools, the sample was stratified according to treatment or comparison status for the grade 4 Balsakhi Program, as well as gender, language of instruction of the school, and average math test scores in the posttest in the previous year. Thus, in the final sample for the study, fifty-five schools received the CAL Program (Group A1B1), and fifty-six served as the comparison group (Group A2B2). The program was continued in

2003–2004, after switching the treatment and comparison groups. Table I summarizes the allocation of schools across different groups in the program.

III.B. Sample: Mumbai

To ensure the results from the Vadodara study would be generalizable, the Balsakhi Program was also evaluated in Mumbai in 2001–2002 and 2002–2003. We selected one ward (the L Ward) to implement a design similar to the design in Vadodara. In total, seventy-seven schools were included in the study. After stratification by pretest score and language of instruction, half the schools were randomly selected to receive a balsakhi in grade 3 (Group C, see Table I), and half the schools were randomly selected to receive a balsakhi in grade 2 (Group D). (Grade 2 students were not included in the study). In 2002–2003, we expanded the study to include students in grade 4. As in Vadodara, children kept their treatment assignment status as they moved from grade 2 to 3 (or 3 to 4).

In the second year of the study, the Mumbai program experienced some administrative difficulties. For various reasons, only two-thirds of the schools assigned balsakhis actually received them. Nevertheless, all children were tested, regardless of whether or not they participated in the program. Throughout the paper, the schools that were assigned balsakhis but did not get them are included in the "intention to treat" group. The regression analysis then adjusts the estimates for the fraction of the treatment group that was effectively treated by using the initial assignment as an instrument for treatment.

III.C. Outcomes

The main outcome of interest is whether the interventions resulted in any improvement in learning levels. Learning was measured in both cities using annual pretests given during the first few weeks of the school year and posttests given at the end of the term.[5]

5. The pretest was administered in July approximately two to three weeks after the official opening of the school in mid-June to ensure that enrollment had stabilized. The one exception was Mumbai year 1: the pretest was administered in late September and early October. The posttest was administered at the end of the academic year, in late March and early April (schools close in mid-April). In addition, in Vadodara, midtests were conducted halfway through the year. Results from these midtests are reported in Banerjee et al. [2005]. They are consistent with the posttest results presented here.

The test covered the basic competencies taught in grades 1–4 and was administered in the school's language of instruction. In what follows, all scores are normalized relative to the distribution of the pretest score in the comparison group in each city, grade, and year.[6]

Differential attrition between the treatment and comparison groups could potentially bias the results. For example, if weak children were less likely to drop out when they benefited from a balsakhi, this could bias the program effect downwards. To minimize attrition, the testing team returned to the schools multiple times, and children who still failed to appear were tracked down at home and, if found, were administered the same test. Table 6 in Banerjee et al. [2005] shows that, except in Vadodara in year 1 (when a number of children left for the countryside due to the major communal riots), attrition was very low. Moreover, in all cases, it was similar in treatment and comparison schools.[7] Furthermore, the pretest scores of children who left the sample were similar in treatment and comparison groups, suggesting that the factors leading to attrition were the same in both groups. These two facts together suggest that attrition is unlikely to bias the results we present below.

Columns (1)–(3) in Table II show the pretest scores' descriptive statistics in the different treatment groups (to save space, the basic descriptive statistics are presented pooling both grades when relevant—the results are very similar in each grade). Columns (1)–(3) give scores for all children present for the pretest, while columns (4)–(6) give scores for children who were present for the pretest and post-test. (Attrition is discussed in the next section.) The randomization appears to have been successful: with the exception of the CAL Program in year 3 in Vadodara, none of the differences between the treatment and comparison groups prior to the implementation of the program are statistically distinguishable from zero. The point estimates are also

6. Scores are normalized for each grade, year, and city, such that the mean and standard deviation of the comparison group in the pretest is zero and one, respectively. (We subtract the mean of the control group in the pretest and divide by the standard deviations.)

7. For the Balsakhi Program, attrition was 17 and 18 percent, respectively, in the comparison and treatment groups in Vadodara in year 1, 4 percent in both the treatment and the comparison group in Vadodara in year 2. In Mumbai it was 7 and 7.5 percent, respectively, in the treatment and comparison groups in year 1, and 7.7 and 7.3 percent, respectively, in year 2. For the CAL Program, the attrition was 3.8 and 3.4 percent, respectively, in year 1 and 7.3 and 6.9 percent in year 2.

TABLE II
TEST SCORE SUMMARY STATISTICS FOR BALSAKHI AND CAL PROGRAMS

	Pretest			Posttest		
	Treatment	Comparison	Difference	Treatment	Comparison	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
A. Balsakhi: Vadodara						
Year 1 (grades 3 and 4)						
Math	−0.007	0.000	−0.007	0.348	0.171	0.177
			(0.059)			(0.070)
Language	0.025	0.000	0.025	0.794	0.667	0.127
			(0.061)			(0.076)
Year 2 (grades 3 and 4)						
Math	0.046	0.000	0.046	1.447	1.046	0.401
			(0.053)			(0.078)
Language	0.055	0.000	0.055	1.081	0.797	0.285
			(0.058)			(0.071)
B. Balsakhi: Mumbai						
Year 1 (grade 3)						
Math	0.002	0.000	0.002	0.383	0.227	0.156
			(0.108)			(0.126)
Language	0.100	0.000	0.100	0.359	0.210	0.149
			(0.108)			(0.102)
Year 2 (grades 3 and 4)						
Math	−0.005	0.000	−0.005	1.237	1.034	0.203
			(0.058)			(0.107)
Language	0.056	0.000	0.056	0.761	0.686	0.075
			(0.054)			(0.061)

TABLE II
(CONTINUED)

	Pretest			Posttest		
	Treatment	Comparison	Difference	Treatment	Comparison	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
C. Computer-assisted learning: Vadodara						
Year 2 (grade 4)						
Math	−0.054	0.000	−0.054	1.129	0.810	0.319
			(0.076)			(0.087)
Language	−0.009	0.000	−0.009	0.719	0.709	0.010
			(0.083)			(0.093)
Year 3 (grade 4)						
Math	0.125	0.000	0.125	0.813	0.232	0.581
			(0.073)			(0.089)
Language	0.116	0.000	0.116	0.118	0.014	0.104
			(0.079)			(0.080)

Notes: This table gives the mean normalized test score for pretest (given at the beginning of the academic year) and posttest (given at the end of the academic year) for treatment and comparison students. Columns (1)–(3) include all children who were present for the pretest. Columns (4)–(6) give the scores for children who were present for the pretest and posttest. Standard errors of the difference, corrected for clustering at the school-grade level, are given in parentheses. The normalized test score is obtained by subtracting the mean pretest score of the comparison group and dividing by the standard deviation of the scores of the pretest comparison group.

very small, with each difference less than a tenth of a standard deviation.

The raw scores and the percentage of children correctly answering the questions relating to the curriculum in each grade (presented in Banerjee et al. [2005]) give an idea of how little these children actually know, particularly in Vadodara. Only 19.5 percent of third grade children in Vadodara and 33.7 percent in Mumbai pass the grade 1 competencies (number recognition, counting and one-digit addition and subtraction) in math. The results are more encouraging in verbal competencies: 20.9 percent of the grade 3 children pass the grade 1 competencies in Vadodara (reading a single word, choosing the right spelling among different possible spellings for a word), and 83.7 percent do so in Mumbai. The baseline achievement level is much higher in Mumbai, where students are less poor than in Vadodara, and schools have better facilities.

Another outcome of interest is attendance and dropout rates. These were collected by Pratham employees who made randomly timed visits to each classrooms every week to take attendance with a roll call. Analysis of this data [Banerjee et al. 2005] demonstrate that both of the programs we evaluate had no discernible effect on attendance or drop out. As a result, we focus here on changes in test scores.

IV. Short-Term Effects

IV.A. Balsakhi Program

Table II presents the first estimates of the effect of the Balsakhi Program—the simple differences between the posttest scores in the treatment and comparison groups.

The Balsakhi Program appears to be successful: in all years, for both subjects, in both cities, and for all subgroups, the difference in posttest scores between treatment and comparison groups is positive and, in most instances, significant.[8] In Vadodara, in the first year, the difference in posttest scores between treatment and comparison groups was 0.18 standard deviations for math and 0.13 for language. The measured effect is larger in the second year, at 0.40 for math and 0.29 for language. In Mumbai in year

8. All standard errors reported in the paper are adjusted for clustering at the school-grade level, the level of randomization.

1, the effects are 0.16 and 0.15 for math and language, respectively. In year 2, the difference between treatment and comparison groups is smaller in Mumbai than in Vadodara: 0.203 for math and 0.075 for language, the language results being insignificant. (Note that Mumbai year 2 results are "intention to treat" estimates since one-third of the schools in the treatment group did not get a balsakhi. (The "treatment on the treated" estimates will be presented in the next table.)

Because test scores have a strong persistent component, the precision of the estimated program effect can be increased substantially by controlling for a child's pretest score. Since the randomization appears to have been successful and attrition was low in both the treatment and comparison groups, the point estimates should be similar to the simple differences in these two specifications, but the confidence interval around these point estimates should be much tighter.

Table III presents the results, for various years, cities, and grades from a specification which regresses the change in a student's test score (post-test score minus pretest score) on the treatment status of the child's school-grade, controlling for the pretest score of child i in grade g and school j:

$$y_{igjPOST} - y_{igjPRE} = \lambda + \delta D_{jg} + \theta y_{igjPRE} + \epsilon_{igjPOST}, \tag{1}$$

where D_{jg} is a dummy equal to one if the school received a balsakhi in the child's grade g, and 0 otherwise.[9] This specification asks whether children improved more relative to what would have been expected based on their pretest score in treatment schools than in comparison schools. For all years and samples, except Mumbai in year 2, (1) is estimated with OLS. However, for Mumbai in year 2 (and when both cities are pooled), to account for the fact that not all schools actually received a balsakhi, (1) is estimated by two stage least squares, instrumenting for actual treatment status of the school-grade ("did the school actually get a balsakhi for that grade?") with a dummy for intention to treat.

In accordance with the simple difference results, these estimates suggest a substantial treatment effect. Pooling both cities and grades together (in the first two rows of Table III), the impact of the program on overall scores was 0.14 standard deviations overall in

9. In Banerjee et al. [2005], we also present a difference in difference specification, which gives very similar results. Estimating (1) without controlling for pretest score also gives very similar results.

TABLE III
ESTIMATES OF THE IMPACT OF THE BALSAKHI PROGRAM, BY CITY AND SAMPLE

	Number of observations	Dependent variable: test score improvement (posttest − pretest)		
		Math	Language	Total
	(1)	(2)	(3)	(4)
A: Pooling grades and locations				
Mumbai and Vadodara together year 1	12,855	0.182	0.076	0.138
		(0.046)	(0.056)	(0.047)
Mumbai and Vadodara together year 2	21,936	0.353	0.187	0.284
		(0.069)	(0.050)	(0.060)
B: Pooling both grades				
Vadodara year 1	8,426	0.189	0.109	0.161
		(0.057)	(0.057)	(0.057)
Vadodara year 2	11,950	0.371	0.246	0.331
		(0.073)	(0.061)	(0.070)
Mumbai year 1 (grade 3 only)	4,429	0.161	0.086	0.127
		(0.075)	(0.066)	(0.067)
Mumbai year 2	9,986	0.324	0.069	0.188
		(0.145)	(0.081)	(0.112)
C: Grade 3				
Vadodara year 1	4,230	0.179	0.102	0.152
		(0.086)	(0.085)	(0.085)
Vadodara year 2	5,819	0.418	0.233	0.354
		(0.107)	(0.089)	(0.100)
D: Grade 4				
Vadodara year 1	4,196	0.190	0.114	0.166
		(0.072)	(0.076)	(0.073)
Vadodara year 2	6,131	0.307	0.240	0.289
		(0.078)	(0.068)	(0.074)
E: Two year (2001–2003)				
Mumbai pretest year 1 to posttest year 2	3,188	0.612	0.185	0.407
		(0.141)	(0.094)	(0.106)
Vadodara pretest year 1 to posttest year 2	3,425	0.282	0.181	0.250
		(0.094)	(0.079)	(0.088)

Notes: This table reports the impact of the Balsakhi Program, for different groups and years. Each cell represents a separate regression of test score improvement on a dummy for treatment school, controlling for initial pretest score. Standard errors, clustered at the school-grade level, are given in parentheses. Estimates, which include Mumbai year 2, use intention to treat as an instrument for treatment. Normalized test score gain is the difference between posttest and pretest for Panels A–D and the difference between posttest in year 2 and pretest in year 1 for panel E. The total score is the sum of the normalized math and language scores.

the first year, and 0.28 standard deviations in the second year, both very significant. The impact is bigger in the second year than the first, for both math (0.35 vs. 0.18) and verbal (0.19 vs. 0.08).

Comparing Mumbai and Vadodara, the effects are very similar for math in both years (0.19 in Vadodara vs. 0.16 in Mumbai in year 1, and 0.37 vs. 0.32 in year 2), but in Mumbai, the effects for language are weaker and insignificant in both years (0.09 and 0.07 in year 1 and year 2), while they are significant in both years in Vadodara. The lower impact of language in Mumbai is consistent with the fact observed above, that most children (83.7 percent) in Mumbai already had some basic reading skills and are therefore less in need of a remedial program that targets the most basic competencies in language. In math, where more lag behind, the program was as effective as it was in Vadodara.

For both cities and both subjects, the effects are very similar in grade 3 and grade 4. Results are also very similar when the analysis is conducted separately for girls or boys (results for these two specifications not reported).

Compared to the other educational interventions, this program thus appears to be quite effective in the short-run. The Tennessee STAR experiment, for example, for which class size was reduced by seven to eight children (from twenty-two to about fifteen), improved test scores by about 0.21 standard deviations [Krueger and Whitmore 2001]. The Balsakhi Program improved test scores by 0.27 standard deviations in the second year by using alternative instructors for part of the day. Moreover, the balsakhis were paid less than one tenth the teacher's salary (a starting teacher earned about Rs. 7,500 at the time, while balsakhis were paid between Rs. 500 and Rs. 750), making this a much more affordable policy option than reducing class size (in the STAR experiment, a teacher aid program did not have any effect). In the conclusion we discuss the cost effectiveness of the program.

IV.B. Computer-Assisted Learning

Columns (4)–(6) of the third panel in Table II show the posttest scores for the CAL program. The math test scores are significantly greater in treatment schools than in comparison schools in both years. In year 2, the math post-test score is, on average, 0.32 standard deviations higher in the CAL schools. In year 3, it is 0.58 standard deviations higher, but this does not take into account the fact that pretest scores happened to be

TABLE IV
IMPACT OF THE CAL PROGRAM, BY YEAR

	Number of observations	Dependent variable: Test score improvement (posttest − pretest)		
		Math	Language	Total
	(1)	(2)	(3)	(4)
A: Effect of the CAL program				
Vadodara both years	11,255	0.394	−0.025	0.191
		(0.074)	(0.082)	(0.083)
Vadodara Year 2	5,732	0.347	0.013	0.208
		(0.076)	(0.069)	(0.074)
Vadodara Year 3	5,523	0.475	−0.005	0.225
		(0.068)	(0.042)	(0.051)
B: Balsakhi and CAL program: Main effects and interactions (Vadodara, Year 2)				
CAL	5,732	0.408	0.017	0.242
		(0.087)	(0.084)	(0.087)
Balsakhi	—	0.371	0.229	0.315
		(0.112)	(0.104)	(0.112)
CAL ∗ balsakhi	—	−0.144	−0.020	−0.086
		(0.141)	(0.134)	(0.141)

This table reports the impact of the CAL program. In Panel A, each cell represents a separate regression, of test score gain on a dummy for treatment school, controlling for initial pretest score. In Panel B, each column represents a regression, of test score improvement on a dummy for the CAL program, a dummy for the Balsakhi program, and an interaction term, as well as a control for initial pretest score. Standard errors, clustered at the school-grade level, are given in parentheses. Normalized test score improvement is the difference between posttest and pretest. The total score is the sum of the normalized math and language scores.

already 0.13 higher in the treatment group in year 3 (as shown in column (3)).

Table IV corrects for this initial difference by estimating (1), where the treatment is the participation of the school in the CAL program. The CAL program has a strong effect on math scores (0.35 standard deviations in the first year (year 2) and 0.47 standard deviations in the second year (year 3)). It has no discernible impact on language scores (the point estimates are always very close to zero). This is not surprising, since the software targeted exclusively math skills, although some spillover effects on language skills could have occurred (for example, because the program increased attendance, or because the children got practice in reading instructions, or because the teachers had reallocated time away from math to reading). The effect on the sum of language and math test scores is 0.21 standard deviations in year

2 and 0.23 standard deviations in year 3. Panel B of Table IV compares the Balsakhi and the CAL effects and examines their interactions in year 2 (2002–2003) when they were implemented at the same time using a stratified design. When the two programs are considered in isolation, the CAL has a larger effect on math test scores than the Balsakhi Program (although this difference is not significant) and a smaller effect on overall test scores (although, again, the difference is not significant). The programs appear to have no interaction with each other: the coefficients on the interaction on the math and overall test score are negative and insignificant.

IV.C. Distributional Effects

The Balsakhi Program was primarily intended to help children at the lower end of the ability distribution by providing targeted instruction to them. However, it could still have helped the higher scoring children either because they were assigned to the balsakhi or because they benefited from smaller classes when their classmates were with the balsakhi.

The program could also have, perversely, harmed children at the bottom of the distribution (by sending them to a less-qualified teacher) while benefiting children at the top of the distribution (by removing the laggards or trouble-makers from the classroom). While this could result in an improvement in the average test score, it should probably not be construed as a success of the program. It is therefore important to know who among the children were affected by the program.

Table V (Panel A for Balsakhi, B for CAL) shows the results for the year 2002–2003 (year 2) broken into three groups to measure test score gains for children who scored in the top, middle, and bottom third in the pretest.[10] For the Balsakhi Program, the effect is about twice as large for the bottom third than for the top third (0.47 standard deviations versus 0.23 standard deviations for the total score). The program therefore does seem to have been more beneficial to children who were initially lagging behind. Children in the bottom group were more than twice as likely to be sent to a balsakhi (0.22 versus 0.09). For the CAL Program, the impact is also higher for the bottom third, but the

10. Result by initial levels are similar for year 1, but the probability of assignment to the balsakhi is not available in that year.

TABLE V
SHORT- AND LONGER-RUN IMPACTS OF PROGRAMS, BY INITIAL PRETEST SCORE

		Program effect in year 2				Persistence of program effect			
	Probability of assignment to balsakhi	Math	Language	Total	Number of observations	Math	Language	Total	Number of observations
Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Balsakhi, 2002–2003									
All children	0.313	0.371	0.246	0.331	11,950	0.053	0.033	0.040	9,925
		(0.073)	(0.061)	(0.070)		(0.047)	(0.041)	(0.041)	
Bottom third	0.446	0.469	0.317	0.425	4,053	0.096	0.097	0.103	3,356
		(0.088)	(0.074)	(0.084)		(0.045)	(0.038)	(0.040)	
Middle third	0.341	0.374	0.240	0.339	3,874	0.021	−0.024	0.001	3,226
		(0.082)	(0.069)	(0.080)		(0.056)	(0.054)	(0.052)	
Top third	0.162	0.229	0.174	0.216	4,023	0.015	0.006	0.009	3,343
		(0.076)	(0.076)	(0.077)		(0.069)	(0.062)	(0.061)	
Panel B: CAL, 2002–2003									
All children	—	0.347	0.013	0.208	5,732	0.092	−0.072	0.008	4,688
		(0.076)	(0.069)	(0.074)		(0.045)	(0.048)	(0.045)	
Bottom third	—	0.425	0.086	0.278	1,962	0.107	0.004	0.046	1,586
		(0.106)	(0.089)	(0.102)		(0.046)	(0.047)	(0.046)	
Middle third	—	0.316	0.005	0.183	1,844	0.085	−0.105	−0.015	1,511
		(0.081)	(0.081)	(0.082)		(0.055)	(0.069)	(0.058)	
Top third	—	0.266	−0.033	0.146	1,926	0.073	−0.105	−0.013	1,591
		(0.073)	(0.081)	(0.078)		(0.072)	(0.064)	(0.068)	

This table reports the effects of the Balsakhi and CAL Programs over the short- and medium-term, according to the child's position in the initial pretest score distribution. Column (1) reports the probability of actually being taught by the balsakhi, conditional on being in a treatment school. Each cell in columns (2)–(8) represents a separate regression of test score gain on a dummy for treatment, controlling for initial pretest score. In Panel A, intention to treat is used as an instrument for treatment. Columns (2)–(4) give the one-year program effect, estimated as the difference in normalized test score between the posttest and pretest in year 2 (2002–2003). Columns (6)–(8) give the cumulative effect of each program one year after both interventions had stopped. The dependent variable for these regressions is the difference between an end of year test in year 3 (2003–2004), and the pretest in year 2 (2002–2003). Standard errors, clustered at the school-grade level, are given in parentheses.

difference is not as large (0.42 versus 0.27 standard deviations for math score, for the bottom and top groups, respectively).

V. Longer-Run Impact

An important consideration in the evaluation of educational interventions is whether or not the changes generated by the interventions persist over time and last beyond the period in which the intervention is administered.

To investigate this question, we start by comparing the effect of being exposed one versus two years to the program: if the effects are durable, they should be cumulative. In the last two rows of Table III, we present an estimate of the impact of two years of exposure to the program. These are estimates of the difference between the year 1 (2001–2002) pretest and year 2 (2002–2003) posttest for students that were in the third grade during the 2001–2002 academic year and in grade 4 in 2002–2003.[11] In Mumbai, the effect of two years of treatment (from year 1 pretest score to year 2 posttest score) is substantially larger than that in either individual year (0.60 standard deviations in math, for example, versus 0.40 for year 2 in grade 4). It seems possible that the foundation laid in the first year of the program helped the children benefit from the second year of the program. The same, however, is not true for the two-year effect estimates in Vadodara where the two-year effect is slightly smaller than the one-year effect in the second year of the program (though it is larger than the first year's effect). A possible explanation for this is the riots, which occurred in the second half of year 1 in Vadodara. Almost all of the gains because of the balsakhi in Vadodara in the first year accrued in the first half of the year (these results can be seen from the midtest results, reported in Banerjee et al. [2005]). In fact, test scores significantly declined in the second half of the year for both treatment and control students, many of whom were traumatized and absent, even when the schools re-opened. It is possible that by the time the following academic year began, most of the gains accrued in the first part of year 1 had been lost.

11. Only children who were in grade 3 in year 1 can be exposed for two years. Thus, the two-year effect is estimated using substantially fewer students than the one-year effect. There was also naturally more attrition in this group, as students migrated or dropped out during the summer break between year 1 and 2. (Attrition was 33 percent in both Mumbai and Vadodara, and again the pretest score of children who did not appear in the posttest did not vary by treatment status. Table 6A of Banerjee et al. [2005]).

We then investigate whether the program effect lasts beyond the years during which the children were exposed. In Vadodara, we were able to test all children in grade 4 and 5 at the end of year 3 (2003–2004), when the Balsakhi Program ended (see Table I). At that point, grade 4 students in Group B schools had been exposed to the Balsakhi Program during the previous year, when they were in grade 3; grade 4 students in Group A had not ever been exposed to the Balsakhi Program. Grade 5 students in Group A had been exposed in the previous year, when they were in grade 4, and many had been exposed in year 1, when they were in grade 3. Grade 5 students in Group B, never exposed to the program, serve as the comparison group. Finally, grade 5 students in Group A2B2 were exposed to the CAL program in grade 4, while grade 5 students in Group A1B1 had never been exposed to the CAL Program. We were able to track a substantial fraction of these children. The attrition rate reported in Banerjee et al. [2005] is only 20 percent, both for treatment and comparison children, and the pretest scores of the attritors is similar to that of the nonattritors.

Columns (4)–(6) of Table V estimate a specification similar to equation (1), using the difference between the 2004 posttest and the 2002 pretest as the dependent variable, and controlling for 2002 pretest scores. The size of the effects falls substantially, and, indeed, for the Balsakhi Program, the average effect becomes insignificant. However, the effect for the bottom third of the children, who were most likely to have spent time with the balsakhi and for whom the effect was initially the largest, remains significant and is around 0.10 standard deviations both for math and language. For the CAL Program, the effect on math also falls (to about 0.09 standard deviations for the whole sample) but is still significant on average and for the bottom third.

It is not quite clear how these results should be interpreted. On the one hand, the fact that, one year after both programs, those who benefited the most from them are still 0.10 standard deviations ahead of those who did not is encouraging. They may have learned something that had a lasting impact on their knowledge. On the other hand, the rate of decay over these two years is rapid: if the decay continued at this rate, the intervention would very soon have had no lasting impact. One possible interpretation is that the increase of 0.10 standard deviations corresponds to the "real" impact of the program and that the remainder of the difference was due a transitory increase due to short-term improvement in knowledge (that was subsequently forgotten), improvement in test-taking ability, or a Hawthorne effect (for example, children exposed to the balsakhi or to computers may feel grateful and compelled to exert their best effort while taking the test). Another interpretation could be that any advantage in terms of learning that these children had over the children in the comparison group gets swamped by the churning that inevitably happens as the children grow older. Perhaps the only way to retain the gains is to constantly reinforce new learning—as we saw in Table III, in Mumbai, the gains persist and cumulate when the intervention is sustained. The only way to answer this question would be to continue to follow these children. Unfortunately, this becomes much more difficult once they have left the primary school where they studied during the program. We, nevertheless, do intend to track them down in a few years to study their long-term cognitive abilities as well as education and labor market outcomes.

It is difficult to compare these results to other evaluations of education programs in developing countries because very few track down children one year after they stopped being exposed to the program. Two notable exceptions are Glewwe, Ilias, and Kremer [2003] and Kremer, Miguel, and Thornton [2007]. Glewwe, Ilias, and Kremer [2003] evaluate the effects of test-score-based incentives for teachers and found that in the short-term such incentives prompted teachers to provide more test preparation sessions, though their effort level did not change in any other observable dimensions. This teacher effort increased test scores initially, but these increases were not sustained two years after the program. Kremer, Miguel, and Thornton [2007] look at the longer-term effects of test-score-based scholarships for girls. They find that the program caused girls' test scores to increase by about 0.28 standard deviations in one of the districts covered by their study in the year in which the girls received the treatment, and that this effect persisted one year after the end of the program. However, the initial impact on boys (which was almost as large as that for girls) decayed. Taking these results together, a clear implication for future studies is that we need to better understand what makes program effects durable.

VI. Inside the Box: Direct and Indirect Effects

The effects of the Balsakhi Program, reported above, are the effects of having been assigned to a classroom that was included

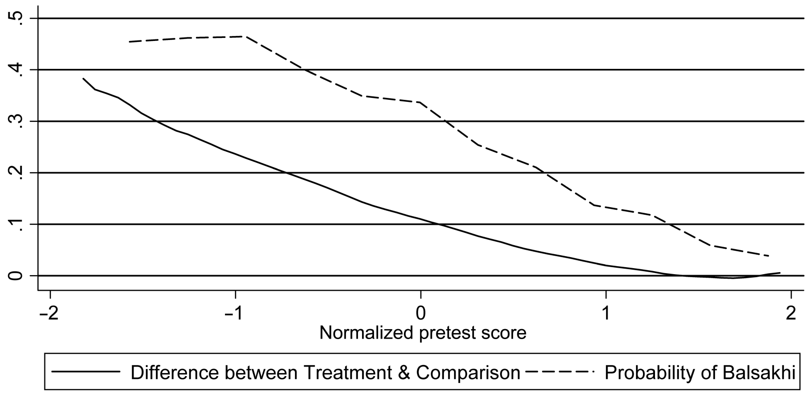

FIGURE I

Program Effect and Assignment Probability as a Function of Pretest Score

Note: The dashed line presents the probability a child is assigned to a balsakhi as a function of her place in the pretest score distribution. The solid line presents the difference in test score gains between children in treatment and comparison groups as a function of their place in the pretest score distribution. The values are computed using locally weighted regressions with a bandwidth of 1.5.

in the Balsakhi Program. As such, it conflates two effects: The program potentially had a direct impact on the children who were assigned to work with the balsakhi. It also could have had an indirect impact on the children who stayed behind in the classroom, both through a reduction in the number of students in the class (a class-size effect) and by removing the weaker children from the room, which could change classroom dynamics (a peer effect).

As we saw earlier, poor initial scorers, who registered the largest gains, were also most likely to be sent to the balsakhi. Figure I plots the difference in test-score gain between treatment and comparison students (the solid line) and the probability of a treatment child being sent to the balsakhi in year 2 (the dashed line) as a function of the initial pretest scores.[12] The test-score gains appear to track closely the probability of assignment to the balsakhi. This suggests that the effect of the program may have been mainly due to children who were sent to the balsakhi, rather than to spillover effects on the other ones.

12. Using a Fan locally weighted regression with a bandwidth of 1.5.

VI.A. Statistical Framework

The ideal experiment to separate the direct and indirect effects of remedial education would have been to identify the children who would have been assigned to work with the balsakhi in all schools, before randomly assigning the schools to treatment and comparison groups. The balsakhi effect could then have been estimated by comparing children designated for the balsakhi in the treatment group with their peers in 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 feasible in this setting since teachers were not prepared to assign the children in the abstract without knowing whether or not they were going to get a balsakhi.

To disentangle these two effects in the absence of this experiment, we use the predicted probability of a child being assigned to the balsakhi in treatment schools as an instrument for actual assignment.

We start by predicting a child's assignment as a flexible function of his or her score in the pretest score distribution:[13]

$$(2) \quad P_{ijg} = (\pi_0 + \pi_1 y_{ijgPRE} + \pi_2 y^2_{ijgPRE} + \pi_3 y^3_{ijgPRE} + \pi_4 y^4_{ijgPRE}) * D_{jg} + \omega_{ijg},$$

where P_{ijg} is a dummy indicating that the child was assigned to the program (i.e., worked with the balsakhi), y_{ijgPRE} is the child's pretest score, and D_{jg} is the dummy defined above, which is equal to one if school j received a balsakhi in the child's grade g, and zero otherwise.

Denote by M_{ijg} the vector $[1, y_{ijgPRE}, y^2_{ijgPRE}, y^3_{ijgPRE}, y^4_{ijgPRE}]$.

We then estimate how the treatment effect varies as a function of the same variables:

$$(3) \qquad y_{ijgPOST} - y_{ijgPRE} = M_{ijg}\lambda + (D_{jg} * M_{ijg})\mu + \epsilon_{ijg}.$$

Equations (2) and (3) form the first stage and the reduced form, respectively, of the following structural form equation:

$$(4) \qquad y_{ijgPOST} - y_{ijgPRE} = \gamma D_{jg} + \tau P_{ijg} + M_{ijg}\alpha + \epsilon_{ijg},$$

13. The results are not sensitive to the number of polynomial terms in pretest scores that we include, i.e., it does not matter if we exclude the fourth- or third- or second-order terms. As we will see later, including more than one term allows us to test the hypothesis that the balsakhi treatment effect does not depend on initial test score.

which we then estimate with an IV regression using M_{ijg}, D_{jg}, and $D_{jg} * M_{ijg}$ as instruments. The coefficients of interest are γ, which gives the impact of being in a balsakhi school but not being assigned to the balsakhi (the indirect effect), and τ, which gives us the impact of working with the balsakhi, over and above the effect of being in a balsakhi school (τ is the direct effect).

This strategy relies on the assumption that the indirect treatment effect of the program (γ) does not vary with the child's score in the initial test score distribution (i.e., that $D_{jg} * M_{ijg}$ can be excluded from the structural equation). To see this, assume, for example, that the indirect treatment effect declined with initial test scores in a way that exactly tracked how the assignment probability changes with the test score. In that case we would mistakenly attribute this declining pattern to the direct effect.

In (4) we have, in addition, assumed that the direct effect does not depend on the child's test score: this assumption simplifies the exposition but is not needed for identification since we have four excluded instruments ($D_{jg} * y_{ijgPRe}$, $D_{jg} * y^2_{ijgPRe}$, $D_{jg} * y^3_{ijgPRe}$, and $D_{jg} * y^4_{ijgPRe}$); we could therefore in principle estimate four parameters rather than one. The four instruments allow us to test this assumption: if the direct effect is constant, (2), (3), and (4) imply that the ratio μ_k/π_k for $k > 0$ (where μ_k is the coefficient on $D_{jg} * Q^k_{ij}$) should all be equal to τ, which can be directly tested with an overidentification test. Note that these equations also imply that if, in addition, γ is zero, the reduced form effect will be proportional to the probability of assignment to the balsakhi, which is what Figure I appears to indicate.

VI.B. Results

In Table VI, we present instrumental variables estimates of the direct and indirect impact of being in a balsakhi group, using the strategy described earlier. The last lines in the table show the F-statistic for the excluded interactions used as instruments, which are jointly highly significant, and the p-value for the overidentification test described in the last paragraph of the previous subsection.[14]

Based on these results, we cannot reject the hypothesis that

14. To save space, we do not report the coefficients from the first stage regression, which is graphically presented in Figure I.

TABLE VI
INSTRUMENTAL VARIABLES ESTIMATES OF DIRECT AND INDIRECT EFFECTS OF PROGRAM

	Dependent variable: Test score improvement (posttest − pretest)		
	Mumbai	Vadodara	Both
	(1)	(2)	(3)
Balsakhi school (γ)	−0.029	0.133	0.056
	(0.085)	(0.106)	(0.068)
Child taught by balsakhi (τ)	0.574	0.614	0.606
	(0.240)	(0.292)	(0.189)
F-stat (first stage)	29.491	78.037	87.586
p-value	0.000	0.000	0.000
Over Id Test: p-value	0.598	0.477	0.476

Table VI presents instrumental variables estimates of the direct (γ) and indirect (τ) effect being in a treatment school. Each column represents a regression. The dependent variable is improvement in normalized test scores; regressions include a control for initial pretest score. Standard errors, corrected for clustering at the school-grade level, are given in parentheses. The F-statistic and p-value from the first stage regression are reported below the regression results. The first stage is presented graphically in Figure 1. The final line reports the p-value from a test of the identifying assumption.

being in a balsakhi school has no effect for children who were not themselves sent to the balsakhi.[15] The effect of the program appears concentrated on children who indeed worked with the balsakhi. The effect on the children sent to the balsakhi is large: they gain 0.6 standard deviations in overall test scores (which is over half of the test-score gain a comparison child realizes from one year of schooling). The overidentification test indicates that we cannot reject the hypothesis 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 assigned to the balsakhi group.

Banerjee et al. [2005] describe and implement a second strategy for separating direct and indirect effects, which exploits the discontinuity in the assignment: students ranked in the bottom twenty of their class are much more likely to be assigned to a balsakhi than those ranked above the bottom twenty. These es-

15. Note, however, that the 95 percent confidence interval of that effect ranges from −0.076 to 0.189. The top of that range is similar to estimates of the class size effects that have been estimated in other contexts.

timates confirm the results reported above: We cannot reject the hypothesis that the program had no effect on children who were not sent to the balsakhi, and while the point estimates of the direct effect are larger than what we report in Table VI (close to one standard deviation), we cannot statistically distinguish them from each other.

VII. Conclusion

This paper reports the results of the impact evaluations of a remedial education and a computer-assisted learning program. Evaluations conducted in two cities over two years suggest that both are effective programs: the test scores of children whose schools benefited from the remedial education program improved by 0.14 standard deviations in the first year and by 0.28 in the second year. We also estimate that children who were directly affected by this program improved their test scores by 0.6 standard deviations in the second year, while children remaining in the regular classroom did not benefit. The computer-assisted learning program was also very effective, increasing math scores by 0.36 standard deviations the first year and by 0.54 standard deviations the second year.

Some may be puzzled by the effectiveness of these two programs and the lack of spillovers of the Balsakhi Program to the other children given that the balsakhis have less training than the formal teachers and that Computer-Assisted Learning Programs have not been shown to be effective in developed country settings. We see two plausible explanations. First, teachers teach to the prescribed curriculum and may not take time to help students who are behind catch up, ending up being completely ineffective for them [Banerji 2000]. Second, students share a common background with the balsakhis but not with the teachers. Ramachandran et al. [2005] argue that social attitudes and community prejudices may limit teachers' effectiveness and that teachers feel as if "they were doing a big favor by teaching children from erstwhile 'untouchable' communities or very poor migrants." These factors may also help explain the effectiveness of the Computer-Assisted Learning program, which allowed each child to be individually stimulated, irrespective of her current achievement level.

Both programs, the Balsakhi Program in particular, are also remarkably cheap, since the salary of the balsakhi (the main cost of the Balsakhi Program) is only a fraction of a teacher's salary (balsakhis were paid Rs 500–750 per month, or a little over \$10–\$15). Overall, the Balsakhi Program cost is approximately Rs. 107 (\$2.25) per student per year, while the CAL Programs cost approximately Rs 722 (\$15.18) per student per year, including the cost of computers and assuming a five-year depreciation cycle.[16]

In terms of cost for a given improvement in test scores, scaling up the Balsakhi Program would thus be much more cost effective than hiring new teachers (since reducing class size appears to have little or no impact on test scores). It would also be five to seven times more cost effective than expanding the Computer-Assisted Learning Program (which brings about a similar increase in test scores at a much higher cost). Banerjee et al. [2005] estimate the cost per standard deviation improvement of both programs under various assumptions, and compare it to other effective programs evaluated in the developing world. The Balsakhi Program, at a cost of about \$0.67 per standard deviation, is by far the cheapest program evaluated. Providing a full cost benefit analysis of these programs is, however, beyond the scope of this paper, since their long-term effects (on learning and on labor market outcomes) are not known.

Nevertheless, these results suggest that it may be possible to dramatically increase the quality of education in urban India, an encouraging result since a large fraction of Indian children cannot read when they leave school. Both programs are inexpensive and can easily be brought to scale: the remedial education program has already reached tens of thousands of children across India. An important unanswered question, however, given the evidence of decay in the gains a year after the programs end, is whether these effects are only experienced in the short term, or can be sustained several years after the program ends, making a long-lasting difference in these children's lives.

MIT Department of Economics and Abdul Latif Jameel Poverty Action Lab
Harvard Business School
MIT Department of Economics and Abdul Latif Jameel Poverty Action Lab
Columbia University Department of Economics, School of International and Public Affairs, and MIT Abdul Latif Jameel Poverty Action Lab

16. In fact, the computers came at no cost to Pratham, so Pratham's annual cost per student was also actually Rs. 367 (\$7.72) per student. Similar situations may be present in many Indian schools. This makes the CAL Program more attractive, but still less cost-effective than the Balsakhi Program.

References

Angrist, Joshua, and Victor Lavy, "New Evidence on Classroom Computers and Pupil Learning," *The Economic Journal,* CXII (2002), 735–765.
Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying Education: Evidence from Two Randomized Experiments in India," NBER Working Paper: No. 11904, 2005.
Banerjee, Abhijit, Suraj Jacob, and Michael Kremer, "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials," MIT Department of Economics Working Paper, 2004.
Banerji, Rukmini, "Poverty and Primary Schooling: Field Studies from Mumbai and Delhi," *Economic and Political Weekly,* XXIII (2000), 795–802.
Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," NBER Working Paper: No. 9671, 2003.
Glewwe, Paul, and Michael Kremer, "Schools, Teachers, and Education Outcomes in Developing Countries," *Handbook on the Economics of Education.* (New York, NY: Elsevier), forthcoming.
Glewwe, Paul, and Sylvie Moulin, "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya," BREAD Working Paper, Cambridge, MA, 2002.
Glewwe, Paul, and Eric Zitzewitz, "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," *Journal of Development Economics,* LXXIV (2004), 251–268.
Hanushek, Eric A., "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature,* XXIV (1986), 1141–1177.
——, "Interpreting Recent Research on Schooling in Developing Countries," *World Bank Research Observer,* X (1995), 227–246.
Kremer, Michael, Edward Miguel, and Rebecca Thornton, "Incentives to Learn," NBER Working Paper: No. 11904, 2007.
Krueger, Alan, and Cecilia Rouse, "Putting Computerized Instruction to the Test: A Randomized Evaluation of a 'Scientifically-based' Reading Program," *Economics of Education Review,* XXIII (2004), 323–338.
Krueger, Alan, and Diane M. Whitmore, "The Effect of Attending Small Class in Early Grades on College Test-Taking and Middle School Test Results: Evidence from Project STAR," *The Economic Journal,* CXI (2001), 1–28.
Lavy, Victor, and Analia Schlosser, "Targeted Remedial Education for Underperforming Teenagers: Cost and Benefits," *Journal of Labor Economics,* XXIII (2005), 839–874.
Leuven, Edwin, Mikael Lindahl, Hessel Oosterbeek, and Dinand Webbink, "The Effect of Extra Funding for Disadvantaged Pupils on Achievement," IZA Discussion Paper No. 1122, 2004.
Machin, Stephen, Costas Meghir, and Sandra McNally, "Improving Pupil Performance in English Secondary Schools: Excellence in Cities," *Journal of the European Economic Association,* II (2004), 396–405.
Machin, Stephen, Sandra McNally, and Olmo Silva, "New Technology in Schools: Is there a Payoff?" Working Paper, London School of Economics, 2006.
Miguel, Edward, and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica,* LXXII (2004), 159–217.
Pratham Organization, "Annual Status of Education Report," (Pratham Resource Center: Mumbai), 2005.
Ramachandran, Vimala, Madhumita Pal, Sharada Jain, Sunil Shekar, and Jitendra Sharma, "Teacher Motivation in India," Discussion Paper, (Azim Premji Foundation, Bangalore, 2005).
Vermeersch, Christel and Michael Kremer, "School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Evaluation" World Bank Policy Research Working Paper: No. 3523, 2005.
World Bank, *World Development Report 2004: Making Services Work for the Poor,* (New York, NY: Oxford University Press, 2004).