

CASE STUDY 2: REMEDIAL EDUCATION IN INDIA

Why Randomize?



This case study is based on “Remedying Education: Evidence from Two Randomized Experiments in India” (by Abhijit Banerjee, Shawn Cole, Esther Duflo, and Leigh Linden) (Poverty Action Lab Working Paper, 2004)

J-PAL thanks the authors for allowing us to use their paper.

KEY VOCABULARY

Counterfactual: what would have happened to the participants in a program had they not received the intervention. The counterfactual cannot be observed from the treatment group; can only be inferred from the comparison group.

Comparison Group: in an experimental design, a randomly assigned group from the same population that does not receive the intervention that is the subject of evaluation. Participants in the comparison group are used as a standard for comparison against the treated subjects in order to validate the results of the intervention.

Program Impact: estimated by measuring the difference in outcomes between comparison and treatment groups. The true impact of the program is the difference in outcomes between the treatment group and its counterfactual.

Baseline: data describing the characteristics of participants measured across both treatment and comparison groups prior to implementation of intervention.

Endline: data describing the characteristics of participants measured across both treatment and comparison groups after implementation of intervention.

Selection Bias: statistical bias between comparison and treatment groups in which individuals in one group are systematically different from those in the other. These can occur when the treatment and comparison groups are chosen in a non-random fashion so that they differ from each other by one or more factors that may affect the outcome of the study.

Omitted Variable Bias: statistical bias that occurs when certain variables/characteristics (often unobservable), which affect the measured outcome, are omitted from a regression analysis. Because they are not included as controls in the regression, one incorrectly attributes the measured impact solely to the program.

INTRODUCTION

At the United Nations Millennium Summit in September 2000, world leaders declared an ambitious set of objectives under the heading of the Millennium Development Goals. Included in the group of eight targets was universal primary education by 2015. While progress is being made towards this important goal, getting students in school is only the beginning. Poor infrastructure, high teacher absenteeism, limited inputs, and large class sizes reduce the quality of education received, especially by poor and under-qualified students. A study in India found that, of all third and fourth graders in Mumbai public schools, 25% cannot recognize letters and 35% do not recognize basic numbers.

UN reports have singled out both sub-Saharan Africa and South Asia as areas lagging in progress in their educational goals. In both these regions, though access to primary schooling has indeed increased, schools are often overcrowded and lack the resources necessary to effectively educate students. A simple comparison of pupil to teacher ratios from 2000 illustrates the gravity of the problem:

TABLE 1

Region or Country	Pupil Teacher Ratio
G7 nations	16.4
Sub Saharan Africa	45.0 (2001)
--Kenya	30.0
South Asia	42.0 (1999)
--India	40.0

Source: World Development Indicators 2004

Embedded in the problem of large classroom sizes is the high variation in student achievement levels within the same class. Lower performing students require different instruction tailored to their specific needs, and in large classrooms, a teacher cannot effectively instruct the mixed student population. Given the opportunity and resources to divide classrooms into smaller units, one possible way to address the educational needs of lower performing students is to stream these students into a particular class in which the teacher will be fully available to focus on them.

The difficulty of providing good education is further compounded by teacher absenteeism and lack of accountability to local officials due to their protected status as civil servants and state government employees. A recent World Bank-funded random survey of 200 schools found no teaching activity in half of the sample of 200 Indian primary

schools (WDR 2004). There are also problems with the centralized hiring of teachers as public servants in developing countries. The guaranteed wages and benefits of public servants add a significant burden to government’s budgets, and in low and middle income countries, teacher salaries amount to 80-90% of primary education spending (WDR, 2004). The push towards universal primary education combined with already strained budgets has created a crisis in providing an adequate number of trained teachers.

USE OF CONTRACT TEACHERS

This teacher supply problem has led researchers to examine programs involving the decentralized hiring of contract teachers. In the broadest sense, contract teachers are teachers who often (but not always) lack the full qualifications of an official government teacher, but who nonetheless meet a certain set of educational requirements and have usually undergone some training. Instead of being hired by the government as public servants, they are usually hired locally by NGOs or village governments on a contract basis.

This structure creates greater accountability for the contract teachers since the hiring, firing, and renewal decisions are not bound by government service rules. Contract teachers may be in charge of their own class, or they may work in tandem with a regular teacher and provide remedial or supplementary instruction. Generally, contract teachers receive no benefits, and their salary is dependent on their specific role. For example, contract teachers in Cambodia in charge of their own classes received pay equal to regular teachers while in Kenya, contract teachers receive roughly one quarter of a regular teacher’s salary and no benefits. In India, contract teachers’ salaries vary considerably:

TABLE 2

Type of Teacher	Monthly salary
Regular teacher	Rs 5,000
Contract teacher in charge of own class	Rs 900 – Rs 3,000
Contract teacher working part time alongside regular teacher	Rs 200 – Rs 1,000

Source: Para Teachers, DPEP Calling

The savings from the much-reduced salary free up resources that can be used to deal with high pupil teacher ratios and variation in student achievement. In addition, the

decentralized hiring provides local communities with the chance to monitor the attendance and instruction of contract teachers and reward or penalize appropriately.

However, programs involving contract teachers are not without their critics. Some critics point to the lower qualifications, training, or experience of contract teachers as indicators that students in these classes will receive poorer instruction. There is also the possibility of gaming in assignment of government teachers as a result of contract teacher programs. The Cambodian program produced some widely publicized scandals involving preferential placement of official teachers into better locations with contract teachers filling in the deficit at the undesirable posts. Others point to possible interschool tensions between contract teachers and official teachers, with contract teachers resentful of their lower pay and official teachers fearful of replacement by the cheaper contract teachers.

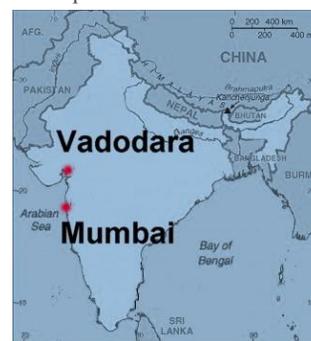
All of these issues must be kept in mind and examined in any study of the effectiveness of using contract teachers in the developing world. In particular we need be concerned about how the overall system adjusts to the presence of these teachers. However, in areas that are facing teacher shortages and very high pupil teacher ratios, contract teachers may have a lot to offer.

THE BALSAKHI PROGRAM: AN EXAMPLE FROM INDIA

In the past 50 years or so, India has made impressive gains in its education system. The number of schools has grown from 223,600 in 1950 to 840,000 in 2004, enrollment has increased from 22.3 million to 155.7 million, and literacy has jumped from 16.6% to 65.4%. Despite these remarkable gains, there are still an estimated 42 million school-age children out of school. Furthermore, 40% of children enrolling in grade one drop out within five years of schooling (Govinda 2004).

Achievement levels, based on testing of basic skills, are equally unimpressive for a significant portion of pupils.

Pratham, a Mumbai-based NGO, with the stated goal “Every child in school...and learning well”, has experimented with



different models to improve education in India. In particular, the Balsakhi Remedial Education Program has great potential due to both its ease to replicate and its low cost (roughly \$5 per child per year). Pratham's decision to expand the Balsakhi program in Vadodara and Mumbai in 2000 presented an opportunity to evaluate the program's effect on student performance with a randomized design.

Presence of Pratham

With support from UNICEF, Pratham was established in Mumbai in 1994 and has since expanded to 39 cities/rural areas in 12 Indian states. As of 2002, Pratham's network of 10,000 workers designed, implemented, and managed programs reaching over 220,000 children. Pratham has established a unique partnership among corporate leaders, government, and Indian communities in which innovative programs implemented by community volunteers and workers enhance education in municipal schools.

The Balsakhi Program

Pratham developed one of its core programs, the Balsakhi Remedial Education Program, in Mumbai in 1994 (expanded to Vadodara in 1999) in response to evidence that a high percentage of children in grades four and five in government schools lacked basic literacy and numeracy skills. In most parts of India, schools have automatic promotions which allow children to advance up to as high as the fourth grade without having to master any of the requisite skills associated with the first four grades. Students who have fallen behind tend to lose interest and drop out or get forced out because the teachers do not want them in class. They also make it harder for other children to learn, since the teacher needs to devote time to remedial lessons.

Literally translated as "friend of child", the "balsakhi" is someone from the local community who has at least completed grade 12. This person is generally female, given the relatively large number of available women with the skills and desire to enter the program. Based on the teacher's aid model in Western schools, the balsakhi for a particular grade works closely for two hours each day (out of a four hour school day) with groups of 15-20 weaker students chosen by the school's instructor. Since there are both morning and afternoon school sessions, she works with two different groups every day for a total of four hours. Pratham has developed a standardized curriculum and provides an initial two-week training session before the school year as well as

ongoing support throughout the year. Pratham is also in charge of hiring and monitoring the balsakhis.

Pratham identifies the following features as key to the design of the Balsakhi program.

- In a small class, the balsakhi can provide more individualized attention, and as a member of the local community, the balsakhi is more familiar with and socially linked to the children.
- Removing children from the classroom benefits non-targeted children by reducing the effective student teacher ratio and by allowing the school instructor to proceed to more advanced topics.
- An effective balsakhi will eventually allow targeted children to return to the mainstream classroom with reinforced basic literacy and numeracy skills.
- The program is easily replicated. Balsakhis are paid roughly \$10/month, are recruited locally, and require relatively little training. Balsakhis also adapt to local space constraints, so there is low overhead and capital costs.
- The balsakhi turnover rate is high (on average a one year stay), so it is unlikely that the program's success depends largely on the ability of a few enthusiastic individuals.
- There is existing evidence that official teachers appreciate the extra help from the balsakhi in reducing class size and helping out with some other basic administrative tasks at the school. Furthermore, because of the high turnover rate and relatively low level of training, there is little threat that they will take over the official teacher's job.

OUTCOME MEASURES: ATTENDANCE AND TEST SCORES

Researchers expected two possible effects of the balsakhi program on schooling: improved attendance and increased test scores. Removing remedial students for part of the day has two potential effects: (1) Lower achieving students are given closer, individualized attention from a local village resident and (2) Students taught by non-balsakhi teachers benefit from a smaller class size for a portion of the day and the ability of teachers to focus on more advanced material, thereby encouraging attendance of higher achieving students.

To test for any educational benefit from the balsakhi in the Vadodara sample, Pratham developed a two-part exam that tested math and language skills separately. This exam covered different skills that the Vadodara Municipal Corporation designated as “compulsory” for each of the grade levels. For example, the math skills test covered topics ranging from basic number recognition, counting, and ordering of single and double-digit numbers to basic addition and word problems. Similar exams were administered in Mumbai.

Tests were administered to all students in the study schools in the grades of interest both at the beginning, middle and end of the school term. This allowed the impact evaluation to focus on improvements rather than the level of performance. (See Discussion Topic 1)

A COMPARISON TO KENYA

Part I: Background of the Kenyan Extra Teacher Program

Akin to India, Kenya faces many of the same constraints in its attempts to achieve universal primary education. There are problems with teacher incentives, costs of educational inputs, overcrowded schools, and low achievement and completion rates. A recent study of Western Kenya found that teachers have absenteeism rates of nearly twenty percent (Glewwe et al, 2003). A 1996 study found primary school completion rates for students who enrolled in the first grade to be 43.3% for girls and 45.1% for boys (Abagi 1997). A 1998 study found that, on average, 4 primary school pupils share one text book (Abagi 1998).

These pre-existing problems have all been exacerbated by the 2003 reintroduction of free universal primary education. With the announcement of the new government policy, enrollment jumped from 6 million to roughly 7.2 million pupils, increasing pupil teacher ratios from 32:1 to 40:1 nearly overnight (Riddell 2003). The Kenyan National Union of Teachers estimates that the country needs an additional 60,000 primary school teachers in addition the existing 175,000, and the country faces a funding crisis (Wax 2003). Even prior to the universal primary education initiative, education accounted for 29% of Kenya’s recurrent expenditure with 55% allocated to primary education and 93% of this used for teacher salaries (Riddell 2003).



Facing overcrowded schools and the inability of the central Kenyan government to provide more teachers, some communities and NGOs have explored contract teachers as a cost-effective means to improve education. As opposed to the civil servants hired by the Ministry of Education (MoE), contract teachers are hired locally by school committees. These contract teachers often have the same qualifications as the civil servants and are out of work due to the squeeze on the national budget. On average, contract teacher receive monthly salaries of 2,000 Kenyan shillings, a fraction of the 8,000 Kenyan shillings price tag of official teachers. The MoE must further provide benefits, amounting to as much as 4,000 Kenyan shillings, to its civil servants.

In January 2005, International Child Support Africa (ICS) initiated a two year program that examines the effect of contract teachers on education in Kenya. Under the program, ICS will give funds to local school committees to hire one extra teacher locally. An initial 2004 pilot conducted on 10 randomly selected schools (out of a 20 school sample) has show great potential, and the full program will be carried out in randomly selected primary schools in the Butere/Mumias and Bungoma Districts in Western Kenya.

Researchers will examine three key components of educational attainment through this project: class size, teacher’s incentives, and peer effects. Smaller class sizes may improve a teacher’s ability to experiment with different teaching methods and will allow the teacher to provide more individualized attention to students. The impact of the smaller class size will be examined through data on promotion rates, drop out rates, test scores, and attendance. The research will also examine how the local hiring of teachers affects teacher’s attendance and performance and how this component affects student outcomes. The local hiring of teachers may provide a cost-effective alternative to the civil servant model if locally hired teachers are indeed as effective as civil servants. Finally, through the randomized design, researchers will also evaluate the effect of peer effects. With an extra teacher, the need to divide students into two classes provides an opportunity to examine how different peers will affect a given student’s performance.

RANDOMIZATION OF THE BALSAKHI PROGRAM

In 2000, Pratham was planning to expand the balsakhi program, which offered an opportunity for evaluation. They

had already been working in schools in Vadodara, a large city in Gujarat, and now planned to move into the remaining 98 municipal schools. Pratham also had expansion plans for Mumbai. Resource constraints precluded the possibility of assigning multiple balsakhis to each school; this limitation, along with the desire to conduct a program evaluation, suggested a randomized experiment.

Researchers were interested in determining what, if any, effect a balsakhi has on students by comparing the change in test scores between schools that received balsakhis (treatment group) and schools that did not receive balsakhis (comparison group). Specifically, researchers had determined to test the balsakhi's possible effect on students in standards three and four in Vadodara and students in standards two and three in Mumbai. In order to stay within the budget, it had been determined that the evaluation would run for two years. Within this period, researchers hoped to examine the effect of the balsakhi on different grades, in different subjects, and over varying spans of instruction.

The research team faced some design problems. First, the nature of the evaluation calls for some schools not to receive balsakhis, but schools would want to participate in the evaluation only if they were to gain something. The exclusionary nature of randomization was therefore politically troubling, but without a comparison group, it would be difficult to attribute any improvement in attendance or achievement to the balsakhi program.

Second, the assignment of schools to groups must be random, so that, on average, the two groups are indistinguishable from each other and represent the general population. Non-randomized group assignment can lead to misleading results. Schools with the lowest initial pre-test scores may have the greatest potential to improve, or conversely, the weaker students in these already low-performing schools may overwhelm the balsakhi's ability even if she were able to help weak students in an average school. In an evaluation design where balsakhis are selectively assigned to the initially weaker performing schools, any results from the data analysis may be due to either the balsakhi or the initial non-randomized assignment of balsakhis. Randomization aims to eliminate this concern since with randomized groups, the evaluation results provide clear evidence that results are due to the balsakhi, and not to any intrinsic difference in treatment and comparison schools.

Discussion Topic 1

Why is a randomized control group necessary?

Some people might consider that a random treatment/comparison design is costly or politically difficult to maintain and to evaluate and that one can compare students who received the program to another set of students who did not receive it. However, randomization is crucial to an unbiased evaluation. The importance of randomization becomes apparent by comparing other plausible, but ultimately misleading, comparisons that researchers could make without the benefit of a randomized control group.

Fallacy 1: Within a grade and school assigned a balsakhi, compare the post-test scores of students who received tutoring to the post-test scores of those who did not.

Figure 2a shows test scores of children that received the balsakhi and of those that did not receive the balsakhi. It suggests that the balsakhi has a negative effect on students' test scores. This conclusion is invalid though. Students sent to a balsakhi are most likely to know less material before any tutoring. This is exactly why they require remedial help. Thus even if the balsakhi helps her tutored students to improve their test scores, this improvement will likely go undetected in this simple differences comparison in a non-randomized framework. This type of analysis does not account for the variation in initial achievement levels of students.

FIGURE 2A

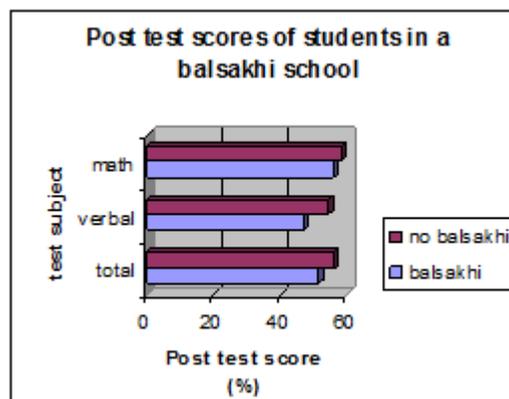
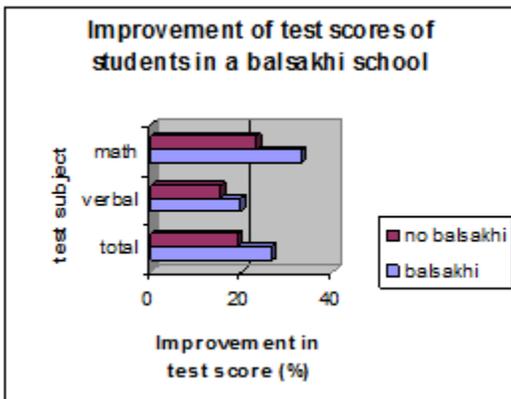


FIGURE 2B



Fallacy 2: Within a grade and school assigned a balsakhi, compare the change in test scores of students who received tutoring to the change of those who did not.

Figure 2b shows that the balsakhi has a large positive effect in improving students' test scores. This conclusion is also misleading. Even though this comparison attempts to detect differences in improvement (as opposed to differences in the post test scores), the result is likely to be biased. Since tutored students generally have such low initial test scores, their improvement is potentially dramatic compared to high performing students. For an extreme example, a student with an initial score of 0% has a much higher margin of improvement than a student with an initial score of 100%. However, the initially low-performing student would almost certainly have a larger improvement than the high-performing student with or without a balsakhi.

Ultimately the following randomization design was judged to be optimal: Researchers first determined at which levels to stratify the sample. Stratification means schools are pre-sorted into groups based on observable characteristics, such as language of instruction. Then from each of these groups, schools are randomly selected to be in the experimental and comparison groups. In the case of language instruction, this assures that there will be an equal number of Gujarati, Hindi, and Marathi language schools in the experimental and comparison groups.

Schools were first stratified by language of instruction and then by student-teacher ratios. For Hindu language schools, there was further need to stratify according to gender of school. Furthermore, in the Mumbai implementation, school were stratified by their pre-test score performance as well.

From these stratified groups, schools were randomly classified as Group A or Group B.

Discussion topic 2

Stratification

Stratification attempts to make sure that the randomly assigned groups are balanced in terms of observable characteristics, such as language of instruction, student-teacher ratios, and gender. However, the concept of randomization entails blindly dividing up a sample into two groups. Why is it that stratification of the sample is necessary prior to randomization?

The Mumbai program had an additional level of stratification, pre test score performance. Presumably, researchers feared that even with randomization, the two groups could have appreciable differences in their pre test scores. Given that this is a concern, why was such stratification not used in the Vadodara sample? Consider some of the pros and cons with stratification and when stratification is not necessary.

CONFIRMING RANDOMIZATION

After stratification and randomization, the research team still had to confirm that Group A and Group B were well balanced, specifically that schools in one of the groups did not have a disproportionate number of schools with certain characteristics. This was done easily by using summary statistics of the two groups to make sure that the random assignment did not produce one group that had higher pre test scores, on average, or differences in any other school level variables that might bias the effect of the balsakhi.

If differences are found in the two groups, then there are some possible solutions. If the imbalance is discovered prior to the program implementation, it is possible to re-run the randomization until the two groups are indeed adequately similar in their original characteristics. If the problem is discovered after program implementation, then there are some statistical tools you can use to correct the error.

Problems with randomization result from the practical limitations of stratification. When there are many dimensions on which the population varies, the researcher has to choose how he wants the sample stratified, since it will not be possible to stratify along every dimension. The schools in the sample had additional observable characteristics, such as whether they held morning or afternoon sessions, their geographic location, and their Muslim student population.

The researchers chose not to stratify along those dimensions, and thereby ran the risk that the two groups, A and B, would be very different along a particular dimension. There is no perfect answer to the question of how one stratifies to avoid this possibility. A rule of thumb is to stratify by the variables that are most likely to have some biasing effect on the program; subsequently, it is important to confirm that there are not big differences in treatment and control groups across other possible confounding variables.

RANDOMIZATION DESIGN

The design was a modified lottery which randomly assigned group assignment at the class level. Each school had two grades participating in the study, but the assignment of grade 3 in a particular school immediately fixed the assignment of grade 4 in that same school.

Discussion Topic 3

Level of Randomization

With the balsakhi program, researchers could have randomized at a different level. Was randomizing at the class level the right choice? Consider the following designs.

Randomization at the School Level

To save on possible administrative costs, researchers might have decided to simply randomly assigned schools to have balsakhis. In this design, schools in Group A would receive balsakhis for both grades in the first year and no balsakhis the second year. Similarly, schools in Group B would receive no balsakhis in year 1 and balsakhis for both grades in year 2. This would allow researchers to monitor the balsakhis more easily since they would have to scrutinize only half of the schools each year. Would there be drawbacks to this approach? What might they be?

Randomization at the Individual Level

Alternatively, researchers could have randomized on the individual level. Due to the variation in school quality, the highest achieving students at say school X may still have lower levels of achievement than lower achieving students at school Y. Some people might argue that it is more equitable to target balsakhis to the lowest achieving students in the entire sample. This suggests a possible approach to randomization.

Suppose the following scenario. With a pre-test, researchers use a cutoff score to identify the lower-achieving students in

all schools “eligible” to receive balsakhi tutoring. This is the relevant sample of students from which half the students (in each grade level) are chosen to receive balsakhi tutoring. Due to proximity of primary schools in an urban Indian environment, students may sometimes go to a neighboring school for tutoring. Thus, if all the students of one grade of one school are low achieving, then it is possible that all of them have an equal chance of being assigned to the balsakhi. Conversely, if a school has only one or two low-achieving students, only they also have a chance of balsakhi assignment, but would likely travel to another school. How does this design affect the treatment and comparison groups in terms of a class size effect? ...in terms of a balsakhi effect on low achieving students?

In Vadodara for the 2001-02 school year, Group A schools received balsakhis for grade 4; Group B, for grade 3. In 2002-03, Group A received schools for grade 3; Group B, for grade 4. Additionally, in this second year, 25 extra schools entered the study, and these were randomly assigned to one of the groups. The randomization design is shown in the table below:

TABLE 3
Balsakhi assignment by grade and randomized group

	Year 1 (2001-02)		Year 2 (2002-03)	
	3 rd grade	4 th grade	3 rd grade	4 th grade
Vadodara				
Group A	NO	YES	YES	NO
Group B	YES	NO	NO	YES
Mumbai				
	2 nd grade	3 rd grade	3 rd grade	4 th grade
Group A	NO	YES	NO	YES
Group B	YES	NO	YES	NO

Pratham also set up a randomized evaluation of the Balsakhi program in Mumbai in 2001-02 to one of the city’s wards (where Pratham had had some presence before). The L-ward included 62 schools, and the randomization was similar to the one employed in Vadodara. Thus the program was evaluated in two different cities in two different States, which ensures that the program is not so specific that it only works in one particular city environment.

To better understand the thought process that drives a successful randomization design, consider possible designs for

the extra teacher program in Kenya. The box on the next page will aid you in this exercise.

Discussion Topic 4

Alternative Randomization Designs

The randomization design is a unique design, and researchers could have gone with more standard models. Consider how the implementation of one of the following designs would have affected the program either through political feasibility, costs, or analysis.

Lottery: Under a pure lottery model, each grade in each school would be its own “observation”. Through random assignment, it is therefore possible that some schools would receive two balsakhis one year (one each for grades 3 and 4) and some schools would receive no balsakhi one year. However, any grade cohort (within a school) that receives a balsakhi the first year would not receive one the second year. Each school would still receive exactly two balsakhis over the two years though.

Phase In: Under a phase in model, each grade in each school would eventually receive a balsakhi. Rather than varying who gets to participate in the program, the researchers randomly vary the timing of participation in the program. Every grade cohort in every school in the sample will benefit from a balsakhi at some point in the program duration. For example, say originally the study was to include 300 balsakhis employed over the 2 years, implying that there would be 150 schools (300 classes) in the study. Under a phase in model, there would be 100 schools (200 classes). In year 1, 100 classes would randomly receive a balsakhi and in year 2, all 200 classes receive a balsakhi. In addition to the reduction in the school sample size, how else does the phase in model affect the possible comparisons?

A COMPARISON TO KENYA

Part II: Randomization Design of the Extra Teacher Program

The Extra Teacher Program in Kenya addresses questions similar to those investigated by the balsakhi program. The researchers have arranged a randomized design that explores three separate issues of educational attainment:

1. The effect of smaller class sizes
2. The effect of hiring teachers locally

3. The effect of having students of the same (or different) quality in a class

The sample size for control and treatment schools is 330. ICS Africa, the implementing NGO, already operates several other randomly designed education and health evaluation programs in the region. These include AIDS education programs and de-worming pill treatment programs. Most schools in the region have already or are currently receiving the benefits of one or more of these programs through ICS. Because of the widespread involvement in other areas, it is plausible for researchers to have some pure comparison schools in this study if necessary.

The structure of the program is important when considering the evaluation design. The extra teachers will be locally hired and generally equally qualified to the government-hired teachers. Thus, the main effect of having a local teacher will be the teacher’s accountability to the local village school committee, which will have the ability to extend or terminate the contract. Anecdotal evidence from the pilot indicates that, if anything, parents prefer the local teachers to the government teachers.

Currently, the largest classes, by far, are all in school’s first grade. Among pilot schools, the average class size was 94.6 pupils. In schools that receive an extra teacher, the first grade class will be divided into two separate classes; the local and government teacher will each teach one of these newly formed classes.

How should the schools be randomized? Note, that given the very large sample size it is possible to run several different treatment groups designed to look at different questions. Different randomization strategies may be used to answer different questions. Consider some of the following key points:

- At what level should the initial randomization occur?
- Is it necessary to use a phase-in model or an original randomization design as in the balsakhi or will a simple lottery suffice?
- Is it necessary to stratify? If so, by what characteristics?
- In addition to simply *providing* smaller classes, are there different ways to divide students into these classes among schools that will be treated?

- What problems may arise if local teachers (or government teachers) are always teaching a certain class type? Is there a fairness issues in assigning teachers?
- If students are divided strictly by ability, might there be parental opposition? How else could you divide students
- When should teachers be asked to divide the class by ability? Which teachers in the sample should be asked to do this exercise?

Keeping in mind the three main issues that researchers will examine with this program, create a chart which shows your randomization design. Based on your design, what groups will you compare to answer the question of classroom effects? ...of the effectiveness of locally hired teachers? ...of peer effects?

How does the program structure of ETP compare to that of the balsakhi program? Why are researchers able to answer more questions with ETP? How does the randomization design depend on the program structure?

SUMMARY OF RESULTS

Before examining test scores in the balsakhi program, researchers had to be certain that the composition of the randomly assigned classes had not changed due to the program. For example, perhaps classrooms assigned balsakhis would have, on average, lower test scores because low-performing students in schools without balsakhis simply drop out at higher rates. In this scenario, the non-balsakhi schools would appear to perform better when the “performance” is in fact due to differential attrition rates between the two groups. However, careful analysis of attendance data shows that there were no measurable effects on attendance or drop out rates from being a treatment school.

The researchers then examined the change in test scores (post minus pre) for treatment and comparison groups. The data below is representative average test scores for fourth graders in Vadodara for the 2002-2003 school year.

TABLE 3:
Analysis of Grade 4 students in Vadodara
(2002-2003 school year)

	Treatment	Control	Difference
Panel A: Pre test scores of grade 4 students in Vadodara			
Sample size (# of students)	3167	3170	-3
Average Math score (%)	31.1	29.7	1.4
Average Verbal score (%)	34.8	33.2	1.7
Average Total score (%)	33.0	31.4	1.6
Panel B: Post test scores of grade 4 students in Vadodara			
Sample size (# of students)	3003	3007	-4
Average Math score (%)	57.5	49.8	7.6
Average Verbal score (%)	51.7	45.7	6.0
Average Total score (%)	54.6	47.8	6.8
Panel C: Change in Scores from Pre test to Post test			
Change in Math score (%)	26.4	20.1	6.2
Change in Verbal score (%)	16.9	12.5	4.3
Change in Total score (%)	21.6	16.4	5.2

The average balsakhi school student had a score improvement from the pre test to the post test of 26.4% on the math portion of the exam while the average comparison school student had a score improvement of 20.1%. The difference in difference, 6.2%, is indeed statistically significant (see text of the academic paper for the formal model). These results generally held true for both math and verbal, for both grades, and for both cities. The two year effect of the balsakhi is even more substantial than the one year effect.

Even with these encouraging results for the average student in the class, researchers still wanted to examine whether the increase was the result of one (or a combination) of two factors. The average test score for the entire class of a treatment group could increase because the balsakhi did indeed provide the remedial help, thereby increasing test scores of the lower achieving students of the class. Alternatively, the improvement in the pupil teacher ratio, achieved by the removal of the lower achieving students in the class, could provide more effective teaching from the official teacher to the higher achieving students of the class.

A problem in the program design prevented researchers from examining this question without additional statistical manipulation of the data. Though it was known which students in a balsakhi class actually received help from the balsakhi, the researchers had no way of knowing who were the students in the control schools who would have gone to the balsakhi had it been a treatment school. To solve this problem the researchers had to make use of a statistical model and some additional assumptions. The resulting statistical analysis did provide strong evidence that the balsakhi-tutored students (the weaker performing students) benefited the most from having a balsakhi assigned to their class. The analysis also showed that there were no indirect effects on the remaining non-balsakhi-tutored students from the balsakhi reducing the effective class size.

MOVING FORWARD

Using a randomized evaluation of the balsakhi program, researchers found strong evidence of the immediate effects of remedial education in teaching basic competencies to under-achieving students in Vadodara and Mumbai schools. Through the carefully planned randomization design, researchers were able to determine that the program successfully targeted improving the weakest students in the overfilled classrooms. Additionally, researchers found no benefits resulting from either the reduced class sizes or more uniform student achievement levels that resulted from the removal of weaker students to their balsakhi tutorials for part of the day.

The balsakhi program is one cost-effective method of helping students who have fallen behind in their education to catch up to the rest of their class. The extra teacher program in Kenya provides another opportunity to examine key issues in educational attainment, such as classroom size, peer effects, and teacher accountability. The more complex randomization design in the Kenyan program will allow researchers to examine multiple questions with one program.

With the global effort towards universal primary education, considerable strain has been added to developing countries' available resources. Children at the bottom end of the distribution, who have the most to learn, often lose out most in school systems plagued by large classrooms, mixed student populations, teacher shortages, and a lack of basic infrastructure and supplies. As local communities and NGOs continue to explore new solutions to these dilemmas,

randomization is a powerful tool to measure program effectiveness in providing education in the developing world.

REFERENCES

- Abagi, Okwach and George Odipo. "Efficiency of Primary Education in Kenya: Situational Analysis and Implication for Education Reform." Institute of Policy Analysis and Research (IPAR) Discussion Paper No. DP 004/97. September 1997. <http://www.ipar.or.ke/dp4.pdf>. January 15, 2005.
- Abagi, Okwach and J. Olweya. "Achieving Universal Primary Education in Kenya by 2015-Where the Reality Lies: Challenges and Future Strategies." Institute of Policy Analysis and Research (IPAR) Discussion Paper No. DP 016/99. 1999. <http://www.ipar.or.ke/> January 15, 2005.
- Banerjee, Abhijit, Shawn Cole Esther Duflo and Leigh Linden 2003. "Remedying Education: Evidence from Two Randomized Experiments in India." mimeo, MIT
- Benaya, Laban, Esther Duflo, Pascaline Dupas, and Michael Kremer 2004. "Decentralizing teacher's hiring procedures: Results of a pilot program in Western Kenya" Poverty Action Lab Working Paper.
- Duthilleul, Yael. "International Perspectives on Contract Teachers and their Impact on Meeting Education for All: The Cases of Cambodia, India, and Nicaragua – Synthesis Report" Conference on Primary School Contract Teachers. UNESCO. <http://www.unesco.org/iiep/eng/research/basic/PDF/teachers3.pdf> February 3, 2005.
- Glewwe, P., Kremer, M., and Nauman, I. 2003. "Teachers Incentives." mimeo, Harvard University and University of Minnesota.
- Govinda, R. and Y. Josephine. "Para Teachers in India: A Review" National Institute for Educational Planning and Administration, New Delhi. UNESCO. October, 2004 draft.
- "Implementation of the United Nations Millennium Declaration: Report of the Secretary General" 2004. United Nations. <http://www.un.org/millenniumgoals/> January 11, 2005.
- "Para Teachers." DPEP Calling. Volume VI, No.11. December 2000. Government of India, New Delhi.

www.educationforallindia.com/page154.html.
January 9, 2005.

<http://www.pratham.org/ourwork/balsakhi.php>.
January 14, 2005.

Riddell, Abby. "The Introduction of Free Primary Education in Sub-Saharan Africa." mimeo, UNESCO. May 16, 2003.
http://portal.unesco.org/education/en/file_download.php/b717832a27e87ed12bc9b77ada00c22bThe+introduction+of+free+primary+education+in+sub-Saharan+Africa.doc
February 5, 2005.

Wax, Emily. "Too Many Brains' Pack Kenya's Free Schools." Washington Post Foreign Service. October 9, 2003, page A24. www.laborrights.org/press/kenya_100903.htm.
January 23, 2005.

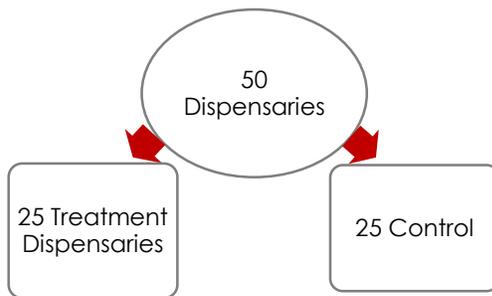
World Development Indicators 2004. The World Bank.

World Development Report 2004. The World Bank.

COMMUNITY-BASED MONITORING OF HEALTHCARE RANDOMIZED EVALUATION

In 2004, researchers conducted a randomized evaluation at 50 dispensaries from nine districts in Uganda to see if community monitoring would improve health worker performance and the impact this might have on health utilization and outcomes.

FIGURE 1



In the 25 randomly selected treatment villages, local NGOs facilitated meetings between the community and their healthcare providers. After community members of all backgrounds had discussed the status of their health services and the steps providers should take to improve health service provision, they then met with health workers to discuss patient rights and provider responsibilities.

The outcome was a shared action plan, or a contract, outlining the agreement between the community and healthcare providers on what needs to be done, how, when and by whom. These meetings were aimed to kick-start the process of community monitoring. Finally, a second set of meetings was held 6 months later to review progress and suggest improvements. More than 150 participants attended a typical village meeting.

THEORY OF CHANGE

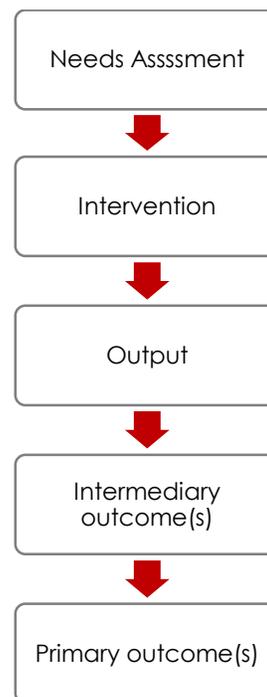
To understand how the programme is intended to have an impact, it is necessary to draw up the Theory of Change from beginning to end.

Discussion Topic 1

Needs assessment and chain of causality

1. What is the need which this intervention is hoping to answer?
2. Using the same framework as used in Session 2, lay out the chain of causality. If you think there could be multiple chains, feel free to draw up more than one.

FIGURE 2



Discussion Topic 2

Assumptions and long-term outcomes

1. What are the assumptions which underlie this chain of causality?
2. Are there any long-term outcomes which you think might be interesting to study

MAPPING FROM TOC TO INDICATORS

Discussion Topic 3

Measuring each step in the chain

1. List several indicators you would use to measure each of 'Output', 'Intermediate Outcomes' and 'Primary Outcomes' in your ToC.
2. What kinds of instruments would you use to collect data on these indicators? Think carefully about issues such as cost and time effectiveness, as well as what sort of instrument will capture that information most accurately.