Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Evaluating Social Programs

Executive Training at the Poverty Action Lab
March 21–25, 2005
MIT, Room E51-372 and Room E51-376

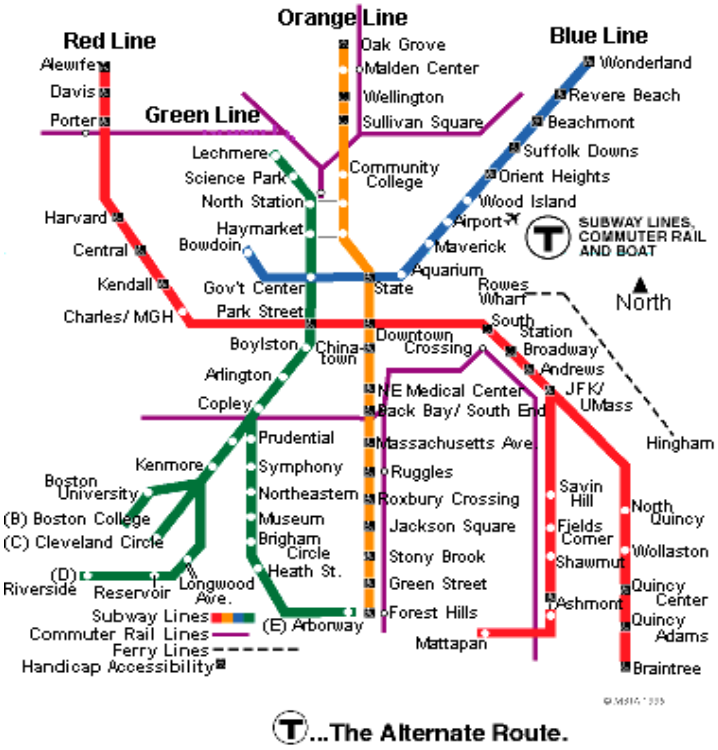
Logistical Information

Hotel Accommodation:

Boston Marriott Cambridge
2 Cambridge Center, (Entrance at Broadway & 3rd Street)
Cambridge, Massachusetts 02142
Phone: 1-617-494-6600
Fax: 1-617-494-0036

Directions from the Boston/Logan Airport:

By Subway: Free shuttle bus service from the airport to the subway system is available on the ground transportation level of the airport. From the *Airport* station located on the blue line, follow the signs labeled *inbound*. Take the inbound blue line train to *Government Center* and switch to a green line train (B, C, D, or E). Take the green line train to the next stop, which is *Park Street* and switch to the red line, going *outbound* towards *Alewife*. Take the outbound red line train to the Kendall/MIT station. Exit the station, and the Marriott will be the building directly behind you.
(Subway fare, $1.25 one-way)



By Cab: Ask your driver to take the Storrow Drive route towards Kendall Square. The address of the Marriott is 2 Cambridge Center.
(Cab fare approximately $30.00)

Driving Directions: Take Callahan Tunnel to 93 North to Exit 26 Cambridge/Storrow Drive. Follow the signs to Storrow Drive and take the Kendall Sq./ Government Center Exit (on left). At the end of the ramp, bear right towards Kendall Square. Go over Longfellow Bridge and the hotel is located two blocks on the left.

Parking:

• On-site parking, fee: $8 hourly, $20 daily
• Valet parking, fee: $25 daily

Restaurants in the area:

Kendall Square

- Au Bon Pain (Cafe)
- Character's Bar & Grill (Marriott)
- Florentina Restaurant (Italian)
- Kendall Hotel
- Kendall Market & Deli
- Legal Seafood
- Polcari's (Italian)
- Rebecca's Café
- Dunkin Donuts

Massachusetts Avenue

- All Asia
- Asgard Pub (Irish bar and grill)
- Chicago Pizza
- Desi Dhaba (Indian)
- Fresco's Grill (breakfast, subs)
- Luna Café
- Miracle of Science (bar & grill)
- Pu Pu Hot Pot (Chinese)
- Royal India
- Sidney's (at the Hotel @ MIT)
- Thailand Café
- Toscanini's ice cream (homemade ice cream)

Main Street

- Bertucci's (pizza, etc.)
- Border India
- Cuchi Cuchi (bar & tapas restaurant)
- Royal East (Chinese)
- Stefani Pizzeria

ATMs and Banks:

Cambridge Trust, Main Street at Kendall Square

Fleet Bank, Main Street at Kendall Square

Post Office:

Located on Main Street at Kendall Square

Central Square:

Central Square is about a 10 minute walk from the hotel (north on Mass Ave.). There you can find drugstores, clothing stores, ethnic restaurants, and both fast food and bistro restaurants as well.

Directions from the Marriott to the Course:

The Poverty Action Lab evaluation course will be held in the MIT Sloan Building. From the Marriott, exit to Main Street, cross the street, and walk left to Wadsworth Street. Turn right onto Wadsworth. Walk just past Amherst Street to your right, and take the stairs located on your left. At the sculpture, bear right towards the Sloan Building. Once inside the building, take the elevators, located just to your left, to the third floor and follow the signs to the classrooms (the course will be held in rooms E51-372 and E51-376).

What to bring to the course:

- Background information or short documents for your ideas on a possible evaluation
- A laptop (if you have one available) to help with your group project work

Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Evaluating Social Programs

Executive Training at the Poverty Action Lab
March 21–25, 2005

Course Participants

Arnab Acharya
Research Economist
The Research Triangle Institute
1615 M St., NW, Suite 740
Washington, DC 20036
aacharya@rti.org

Adedayo Adegoke
Executive Officer/Protocol
Centre for Energy Research and Development
PMB 024, Obafemi Awolowo University
ILE-IFE, Osun/Nigeria 234/220001
aoadegoke@yahoo.com

Livia Benavides
Senior Social Sectors Specialist
World Bank
Av. Alvarez Calderon 185
San Isidro, Lima 27, Peru
lbenavides@worldbank.org

Guush Berhane
Ph.D. Candidate
Wageningen University: Agricultural Economics and Rural Policy Group
Mansholtlaan 10-12
NL-6708 PA Wageningen, The Netherlands
Guush.Berhane@wur.nl

Juan Garrido
Professor
Escuela de Post Grado – Universidad San Ignacio de Loyola
Av. Salaverry 2625
San Isidro, Lima 27, Peru
jgarrido@sil.edu.pe

Philip O'Keefe
Lead Social Protection Specialist
World Bank
14 Martin Place
Sydney, Australia 2000
pokeefe@worldbank.org

William Parienté
Manager, PlaNet Finance
13, rue Dieumegard
93400 Paris St. Ouen - France
Ph.D. Candidate, University of Paris- Sorbonne: Economics
wpariente@planetfinance.org

Sukai Prom-Jackson
Senior Education Specialist
World Bank
1818 H St., NW
Washington, DC 20433
spromjackson@worldbank.org

Frank Rubio
Regional Technical Advisor (Latin America)
Catholic Relief Services
Avenida Vasco Nuñes de Balboa 610 Miraflores
Lima, Peru
frubio@crsperu.org.pe

Luz Magdalena Salas
Consultant
National Planning Department of Colombia
Calle 26 # 13-19, Piso 36
Bogotá, Colombia
lsalas@dnp.gov.co

Patricia Laverley
Operations Evaluation Department
African Development Bank
Angle des trois rues: Avenue du Ghana, Rue Pierre de Coubertin, Rue Hedi Nouira, BP. 323
1002 Tunis Belvedère
Tunis, Tunisia
p.laverley@afdb.org

Sebastián Torres Ledezma
Ph.D. Candidate
University of Leicester: Department of Economics
University Road
Leicester LE1 7RH
United Kingdom
st95@le.ac.uk

Adriana Márquez
Consultant
National Planning Department of Colombia
Calle 26 # 13-19, Piso 36
Bogotá, Colombia
amarquez@dnp.gov.co

Alain Mingat
Lead Education Specialist
World Bank
1818 H St., NW
Washington, DC 20433
amingat@worldbank.org

Nadeem Mohammad
Senior Operations Officer
World Bank
1818 H St., NW
Washington, DC 20433
nmohammad@worldbank.org

Ashwani Muthoo
Senior Evaluation Officer
UN International Fund for Agricultural Development
Via del Serafico 107
Rome, Italy 00142
a.muthoo@ifad.org

Binh Nguyen
Economist
Asian Development Bank
6 ADB Avenue
Mandaluyong, Metro Manila, Philippines 1550
bnguyen@adb.org

Corinne Siaens
Education Economist
World Bank
1818 H St., NW
Washington, DC 20433
csiaens@worldbank.org

Maha Singh
Commissioner
Government of Haryana
21, Staff Road Ambala
Haryana/India 110001
mahasingh@yahoo.com

Jee-Peng Tan
Lead Economist
World Bank
1818 H St., NW
Washington, DC 20433
jtan@worldbank.org

Ajay Tandon
Economist
Asian Development Bank
P.O. Box 789
Manila, Philippines 0980
atandon@adb.org

Inga Tsutskiridze
Program Development/Evaluation Specialist
United States Agency for International Development
20 Telavi Street
Tbilisi, Georgia 0103
itsutskiridze@usaid.gov

Melany Grecia-Viajante
Research Assistant
Center for Agriculture and Rural Development, Inc.
20 M. L. Quezon St., City Subd., San Pablo City
4000 Laguna, Philippines
mbgrecia@yahoo.com

Delia Welsh
Associate Director (Monitoring)
Millennium Challenge Corporation
1000 Wilson Blvd., Suite 1400
Arlington, VA 22209
WelshDC@mcc.gov

Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Evaluating Social Programs

Executive Training at the Poverty Action Lab
March 21–25, 2005
MIT, Room E51-372 and Room E51-376

Course Program

Monday, March 21st
Room E51-372

8:30 AM – 9:30 AM	Continental Breakfast and Welcome
9:30 AM – 11:00 AM	Lecture 1: What is Evaluation? *Lecturer: Sendhil Mullainathan (Harvard University)*
11:00 AM – 12:00 PM	Group Discussion and Casework: Case 1 (Flipcharts)
12:00 PM – 1:30 PM	Lunch
1:30 PM – 3:00 PM	Lecture 2: Why Randomize? *Lecturer: Dan Levy (Mathematica; Harvard University)*
3:00 PM – 4:30 PM	Group Project Work (groups assigned; topics for presentation chosen) *TA available*
4:30 PM – 6:00 PM	Group Discussion and Casework: Case 2 (Balsakhi) *TA available*
6:00 PM – 8:00 PM	Group Dinner

Tuesday, March 22nd
Room E51-372

8:00 AM – 10:30 AM	Continental Breakfast and Group Project Work/Casework: Case 2 (Balsakhi) *TA available from 9:00–*
10:30 AM – 12:00 PM	Lecture 3: How to Randomize? Part I *Lecturer: Abhijit Banerjee (MIT)*

12:00 PM – 1:30 PM	Lunch
1:30 PM – 3:00 PM	Lecture 4: How to Randomize? Part II *Lecturer: Dean Karlan (Princeton University)*
3:00 PM – 4:30 PM	Group Project Work *TA available*
4:30 PM – 6:00 PM	Group Discussion and Casework: Case 3 (Panchayats) *TA available*

Wednesday, March 23rd
Room E51-372

8:00 AM – 10:30 AM	Continental Breakfast and Group Project Work/Casework: Case 3 (Panchayats) *TA available from 9:00–*
10:30 AM – 12:00 PM	Lecture 5: Measurement and Outcomes *Lecturer: Esther Duflo (MIT)*
12:00 PM – 1:30 PM	Lunch
1:30 PM – 3:00 PM	Lecture 6: Sample Size and Data Management *Lecturer: Esther Duflo (MIT)*
3:00 PM – 4:30 PM	Group Project Work *TA available*
4:30 PM – 6:00 PM	Group Discussion and Casework: Case 4 (Deworming) *TA available*

Thursday, March 24th
Room E51-372

8:00 AM – 10:30 AM	Continental Breakfast and Group Project Work/Casework: Case 4 (Deworming) *TA available from 9:00–*
10:30 AM – 12:00 PM	Lecture 7: Managing Threats *Lecturer: Abhijit Banerjee (MIT)*
12:00 PM – 1:00 PM	Lunch

1:00 PM – 2:30 PM	Lecture 8: Analyzing Data *Lecturer: Michael Kremer (Harvard University)*
2:30 PM – 4:30 PM	Finalize Group Project Presentation *TA available*
4:30 PM – 6:00 PM	Group Discussion: Impacting Policy

Friday, March 25th
Room E51-376

8:00 AM – 9:00 AM	Continental Breakfast
9:00 AM – 10:30 AM	Lecture 9: Policy Impact and External Validity *Lecturer: Rachel Glennerster (Poverty Action Lab, MIT)*
10:30 AM – 12:00 PM	Group Presentations
12:00 PM – 1:30 PM	Lunch
1:30 PM – 3:00 PM	Group Presentations
3:00 PM – 4:00 PM	Closing Remarks

Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Evaluating Social Programs
Executive Training at the Poverty Action Lab
March 21–25, 2005

Group Work Instructions

You will be assigned to groups of 4-5 people. We will do our best to ensure that each group includes participants with a range of different experiences but some common areas of interest. You will carry out two types of activities within these groups:

i) casework and discussions;

ii) preparation of group proposal.

Casework and Discussions

Each case covers a specific set of topics which are the subject for the lectures for each day of the course. The cases provide background on one (or in some cases two) specific evaluations which will be referred to in the lectures. In addition, each case includes discussion topics designed to get you thinking about the issues before the lectures. Some of the cases also include exercises for you to complete. You will be provided with Excel files containing these exercises at the start of the "group work" sessions. You will be expected to read the relevant case, go through the discussion topics, and complete the exercises before the related lecture on the case.

Group Proposal

Each group will—throughout the week—work on a proposal for an evaluation on a topic of their choice. Different aspects of evaluation will be covered in the lectures and the casework, and these should be reflected in the group proposal. On Friday, each group will present their proposal and receive comments from the other participants and the lecturers. This is an ideal time to get feedback on an evaluation you may be planning.

The outputs for the project will be:

i) a 20-minute presentation (with an additional 10 minutes for questions and feedback);

ii) a 3-4 page note summarizing the proposal.

Both the presentation and the note should cover the following issues:

i) the objective and rationale of the evaluation—what is the question you are asking and why is it important or interesting?

ii) randomization design—how will the treatment and control groups be determined, and at what level will the randomization take place?

iii) measurement issues—how will you measure whether the program is a success? On what variables will data be collected? How will it be collected? In addition to final outcome measures, will you be collecting data on the mechanism by which the program works? If so, what data will you collect on this?

iv) What magnitude of effect will you be trying to detect? What is the sample size you will be using? Why is this the right sample size?

v) What are the risks to the integrity of the evaluation? How will you seek to minimize these?

vi) How will the data be analyzed?

vii) To what use will you put the results? How will the results impact future policy/programs?

Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Case 1: Flipcharts
Classroom Inputs in Kenyan Schools: An Evaluation of Evaluations

This case study, with kind permission of the authors, is based on "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," (by Paul Glewwe, Sylvie Moulin and Eric Zitzewitz), Journal of Development Economics. (NBER Working Paper 8018, 2000)

Introduction

In Teso and Busia, a neighboring pair of agricultural districts in Kenya, the school boards, in conjunction with researchers from the US, were searching for tools to improve the quality of education. They explored the value of flipcharts as a supplemental aid in teaching the districts' primary school children.

Two separate evaluations were undertaken. The first was a standard regression analysis, taking data on the number of flipcharts in schools on the one hand, and student test score data on the other, to assess whether there was a positive relation between the two. This was called an *observational study* because the researcher/evaluator was an observer who did not decide which schools had flipcharts and which schools did not.

The second evaluation used a different approach. In contrast to the observational study, this was a *randomized experiment*, and was therefore conducted in a way such that the evaluator/researcher had control over the allocation of flipcharts. Specifically, a sample of schools was chosen, and each school was randomly separated into treatment and control groups—where the treatment group received flipcharts and the control group did not. The two groups of schools were later compared against each other to assess the effect of flipcharts on test scores.

These two districts served as the laboratory, not only to evaluate a particular program, but also to evaluate two evaluation methods.

Background

School Quality and Educational Resources

Educational quality in developing countries is typically very low, particularly in rural areas. Resources for education in these communities are also lacking. There is a long standing debate, however, over the policy implications of this correlation. On one side of the debate are those who suggest that more money should be pumped into educational resources such as instructional aids. On the other side there are those who, citing empirical evidence, argue that expenditure on school inputs has been shown to have little affect on school quality (Hanushek, 1995). Both sides tend to agree that inputs vary widely in terms of quality and effectiveness.

From Textbooks to Flipcharts

A study conducted in an agricultural region in the Western Province of Kenya illustrates the complexity of trying to estimate the relationship between school inputs and school quality.

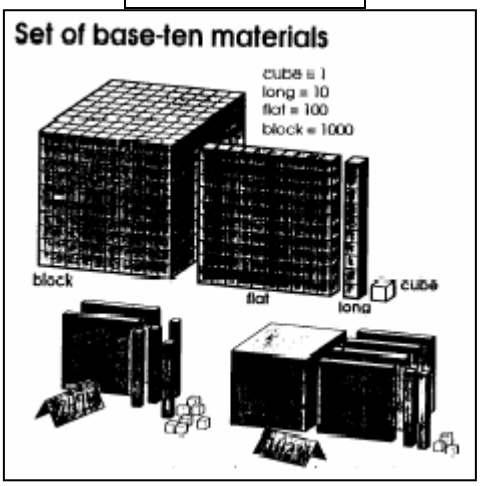
The children in the districts' roughly 300 public schools were deprived of essential teaching and learning materials. Textbooks, for example, were rare. In 8th grade, which typically only the better performing children reach, about 40 percent of students had textbooks in math and English, but 15 percent or less had textbooks in science and other subjects. In lower grades, textbooks were even less common.

To investigate the impact of providing textbooks in these circumstances, the authors had conducted a study in which all students in a randomly selected 25 of the districts' 100 lower-performing schools were given textbooks. They discovered no impact for the bottom 60 percent of the class. They hypothesized that this might reflect the fact that textbooks were written in English—the third language of most children in the region. Perhaps the weaker students had not known enough English to benefit from the textbooks. The authors therefore looked at other, hopefully more effective and less costly, educational inputs. (Glewwe et al, 2000)

There was suggestive evidence from previous research that visual aids would promote learning in many different subjects, such as social studies (Davis, 1968), anatomy (Dwyer, 1970), ecology (Holliday, 1973), and reading (Samuels, 1970). Students more often recall having seen pictures than words or sentences (Shepard, 1967). In addition, learning styles vary across students, so adding visual aids may reach a broader range of students—especially when many students may have difficulty understanding English, the language spoken by the teachers as well as that of the textbooks.

Given the teaching potential of visual aids, the authors decided to conduct an evaluation on the impact of flipcharts in these schools.

Math Flipchart



Outcome Measure:

The process of evaluation requires more than just identifying inputs (in this case, flipcharts). Equally important is deciding what the intended impact is supposed to be and how to measure it. Interested in whether flipcharts improve *competency level* or *learning*, we need a measurement tool that quantifies this. Typically, educational outcomes are measured by

standardized assessment tests. In evaluating educational inputs, the assumption is that any input that affects learning or knowledge should be reflected in test scores.

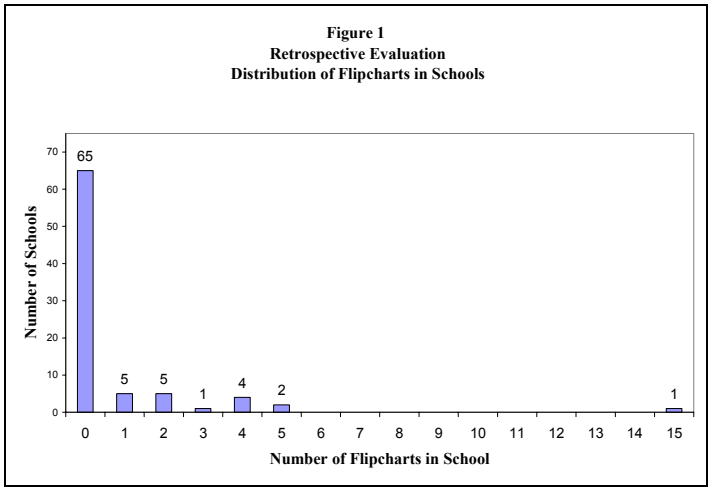
In Kenya, 8th grade is a pivotal year for students. At the end of the year students take an exam, the results of which determine whether they receive the Kenya Certificate of Primary Education (KCPE). Only with this certificate, can students proceed to secondary school. This exam is highly competitive and covers the entire range of subjects taught in school. This exam is typically used to measure the level of knowledge of students in the 8th grade. At the beginning of the school year, and in earlier grades, students take practice exams which cover the same material.

All appropriate tests were used to measure flipcharts' effect on competency level.

Observational Study

Given that flipcharts could already be found in different amounts in different schools, the authors could measure whether those schools with flipcharts fared better than those without, and whether schools with more flipcharts fared better than schools with fewer. And with the exam data available to measure student performance, the stage was set for a standard regression analysis.

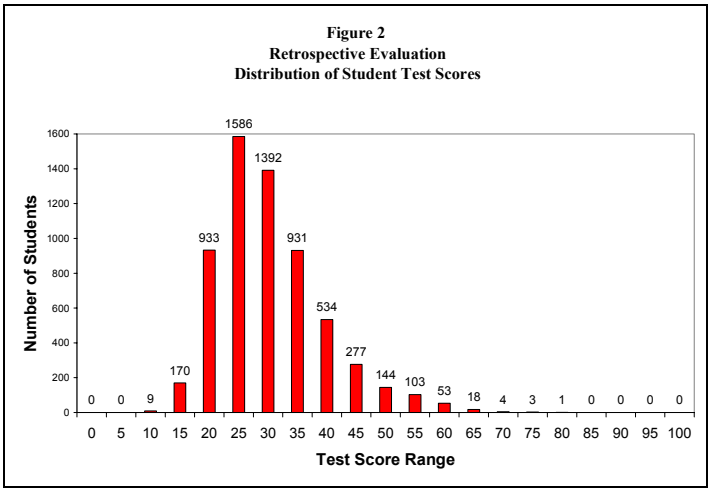
Of the 100 schools in the textbook evaluation, 83 schools had all the relevant data for running this particular regression. Flipcharts were used in three subjects--science/agriculture, math, and home science/business education. Eighteen of the 83 schools had flipcharts, 65 had none at all. And among the schools that had flipcharts, some schools had one flipchart only, and others had up to 15. Figure 1 presents the distribution of flipcharts among the 83 schools:



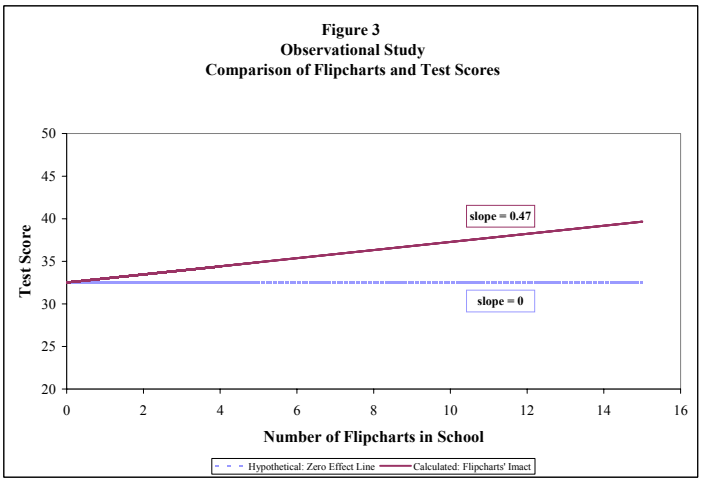
Flipcharts were used in three subjects. The data on the number of flipcharts were not broken down by subject, however. Data were available only on the total number of flipcharts. The authors were therefore unable to relate the specific flipchart subject and

test subject. Instead, they compared the total number of flipcharts to the combined test score in the three relevant subjects.

The exams covered 7 subjects: math, science/agriculture, home science/business education, English, Swahili, geography/history/civics/religion, and arts/crafts/music. Of these seven, the authors only analyzed the first 3 subjects in the initial regressions (the flipchart subjects). Figure 2 presents the distribution of test scores in the relevant 3 subjects among the 83 schools. The scores are scaled to 100:



The average test score is around 36. But with a distribution of test scores, and a distribution of flipchart availability, the question of interest was: how are these two distributions related? Do students from flipchart schools learn more over the course of the year? Figure 3 represents the relation.



The comparison line has a positive slope which illustrates a positive correlation between flipcharts and test scores. This means that students in schools with more flipcharts, in fact, tend to learn more. (The regression line suggests that students with no flipcharts

seem to receive an average score in the low 30s. Students with 5 flipcharts seem to score closer to the mid-to-upper-30s range.) If there were no relationship, we would expect to see a flat, horizontal line (the dotted line).

But how "true" is this relationship? It is very likely that the *exact* slope we predicted is *slightly* off. And perhaps our estimation is very off. Perhaps the true slope is zero. Using standard statistical methods, it can be shown that if the true slope were in fact zero, there is less than a one percent chance we would have observed a slope this positive. This appears to suggest a very high degree of confidence in this result.

This observational study found that an additional flipchart would lead to an increase in test scores of 0.47 points, suggesting that flipcharts were effective (see Table 1 or "slope" in Figure 3). But as can be seen in Table 2 below, there are other differences between schools that have flipcharts and schools that do not, apart from the availability of flipcharts. For example, schools with flipcharts were more likely to have blackboards than schools without flipcharts. So it is possible that the greater availability of blackboards was at least partly responsible for why students in schools with flipcharts did better than students in schools without flipcharts.

Table 1:
Impact of Each Additional Flipchart on Test Scores
Analyizing Flipcharts, Alone,
And Controlling for Other Variables

Test Score Impact from Each Additional:[1]	**Evaluation With**	
	Only Flipcharts	**All Variables**
Flipchart	**0.47**	**0.49**
Pupil Age		-0.27
Teacher Training Level		0.00
Indoor Classroom		14.74
Non-Leaky Roof		-0.91
Desks Per Pupil		-0.01
Blackboard		-3.00
Textbook/Pupil		1.23
Class Size		-0.02

Table 2:
Distribution of Other Variables
Between Schools With Flipcharts
and Schools Without Flipcharts

Other Variables (Averages)[1]	**Schools with:**	
	No Flipcharts	**Flipcharts**
Flipchart	**0**	**4.15**
Pupil Age	**14.31**	14.04
Teacher Training Level	2.05	**2.07**
Indoor Classroom	0.96	**0.98**
Non-Leaky Roof	0.97	**1.00**
Desks Per Pupil	**0.40**	0.37
Blackboards	0.91	**0.95**
Textbooks/Pupil	0.25	0.25
Class Size	32.94	**34.66**

[1] The Variable "Teacher training" level is on a scale between 1 and 6, with 1 being the lowest training possible and 6 being the highest.
"Indoor classroom" can either be 1, which means that the classroom is indoors, or 0, which means class is held outside
Non-Leaky roof is for indoor classrooms with 1 indicating a non-leaky roof and 0 indicating a leaky roof

Fortunately, the researchers had data on availability of blackboards so they were able to account for this difference in the analysis. Column 2 of Table 1 displays the results of their study taking into account ("controlling for" or "holding constant") these other differences between schools with flipcharts and schools without flipcharts. These new results (in which the suggested effect cannot possibly be due to differences in availability of blackboards since these differences have already been explicitly "accounted for") suggest that an additional flipchart leads to an increase in test scores of 0.49 points, also indicating that flipcharts were effective.

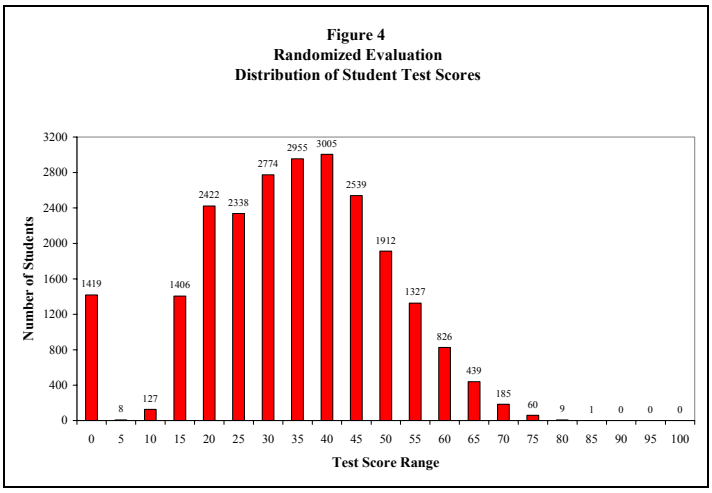
Discussion Topic I: Observational Study
Do you believe the results of this evaluation reflect the true effect of flipcharts on student test scores for this set of schools? Why? Why not?

Randomized Evaluation

For the randomized evaluation, a separate set of schools were used. Of the remaining 200 plus schools in the districts, 178 were eligible (meaning, they had no existing flipcharts). As opposed to those in the prior evaluation, these schools were closer to the median-quality school, and were therefore more representative of the district as a whole.

Half of the schools were randomly selected to receive flipcharts. This group is labeled as the *treatment group*. The other 89 schools acted as the comparison group, or *control group*. Flipcharts were donated by a Dutch NGO, through a local partner, International Child Support Africa (ICS). They included two sets of science charts, one set of charts in math, one in health, and a wall map.

Figure 4 shows the distribution of test scores after 1 year of flipcharts being in place.



Looking at Figure 4, it is clear that, as in the retrospective evaluation, not all students were at the same level at the end of the year. The mean test score for all schools is around 36 points. But how does this distribution break up between treatment and control schools?

This question is answered in Figure 5.

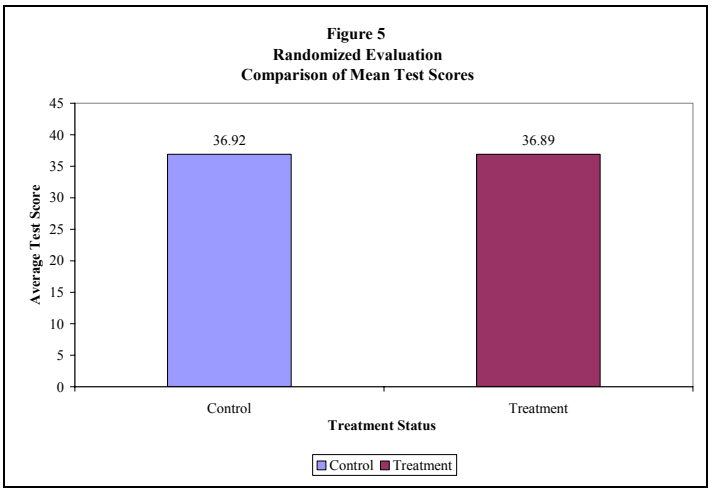


Figure 5 shows the mean test scores in the treatment and control groups. Unlike the prior evaluation, in which the flipchart-to-test score graph had a line with a positive slope, therefore showing a positive relation between the two, in this figure, there doesn't seem to be much difference in test scores at all. Regardless of whether schools had flipcharts or not, their test-scores seemed to be about the same. This suggests that flipcharts had zero impact on child performance.

Discussion Topic II: Randomized Evaluation
Do you believe the results of this evaluation reflect the true effect of flipcharts on student test scores for this set of schools? Why? Why not?

Conclusion

The purpose of this case study is not to evaluate flipcharts. The purpose of this case study is to *evaluate evaluation methods*. Clearly, the two evaluations used to measure flipcharts are not equal.

Discussion Topic III: Comparison of Evaluations
Why do you think the results from the two evaluations are different?

References

Davis, O. L., Jr. 1968. "Effectiveness of Using Graphic Illustrations with Social Studies Textual Materials. Final Report." Kent State University, Ohio..

Dwyer, Francis M., Jr. 1970. "Exploratory Studies in the Effectiveness of Visual Illustrations." AV Communication Review, 235-49.

Glewwe, Paul, Michael Kremer and Sylvie Moulin. 2000. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Policy Research Group. The World Bank.

Hanushek, Eric. 1995. "Interpreting Recent Research on Schooling in Developing Countries." World Bank Research Observer, August, pp. 227-46.

Holliday, William G. 1973. "A Study of the Effects of Verbal and Adjunct Pictorial Information in Science Instruction." Mimeo, Ohio State University.

Samuels, S.J. 1970. "Effects of Pictures on Learning-To-Read, Comprehension, and Attitude." Review of Educational Research, Vol. 16, 397-407.

Shepard, Roger. 1967. "Recognition Memory for Words, Sentences, and Pictures." Journal of Verbal Learning and Verbal Behavior, Vol. 6.

Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

Case 2: Remedial Education in India: A Randomized Evaluation of the "Balsakhi Program"

This case study, with kind permission of the authors, 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)

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:

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 Dev elopement 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:

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 – 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 establish 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) Non-balsakhi taught students 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.

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 role 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.

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 I)

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 I: 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.

Fig 2a

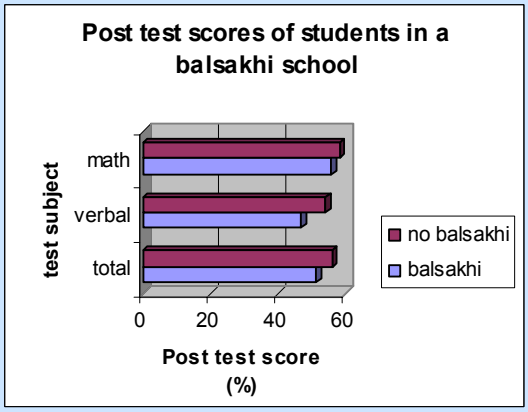
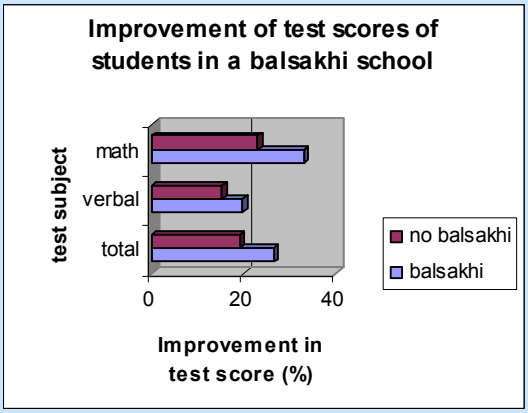


Fig 2b



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 II: 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.

The exercise that accompanies this case should help you think through some of these issues.

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 a significant impact on the program or the outcome variables; 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 groups 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 III: 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 for this. 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 1: Balsakhi assignment by grade and randomized group

	Year 1 (2001-02)		Year 2 (2002-03)	
Vadodara	**3rd grade**	**4th grade**	**3rd grade**	**4th grade**
Group A	NO	YES	YES	NO
Group B	YES	NO	NO	YES
Mumbai	**2nd grade**	**3rd grade**	**3rd grade**	**4th 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 IV: Alternative Randomization Designs

The randomized designwe have used 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 most schools are receiving some kind of assistance from ICS, it is plausible for researchers to have some pure comparison schools (in the sense that they do not receive extra teachers) 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 a 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 than 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 2: 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 (number 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 (number 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.educationforallinindia.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.

Exercise for Case 2: The Importance of Being Stratified

In the balsakhi case, you were introduced to the concept of stratification. This exercise will demonstrate the importance of stratifying data prior to randomization to be assured that the treatment and comparison groups are generally indistinguishable by relevant characteristics. Without this balance of characteristics across the treatment and comparison groups, any outcome differences detected by the evaluation may be attributed to these pre-existing differences of the two groups.

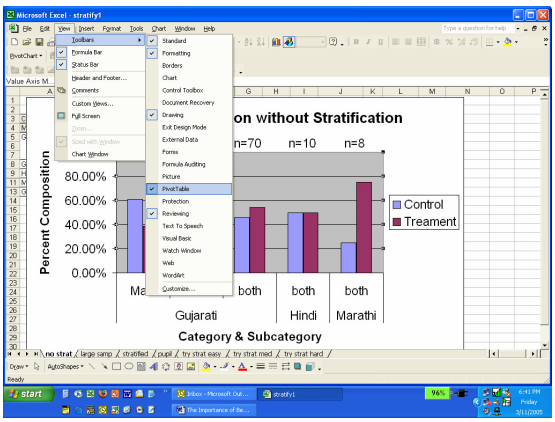
Section I: Random but Biased

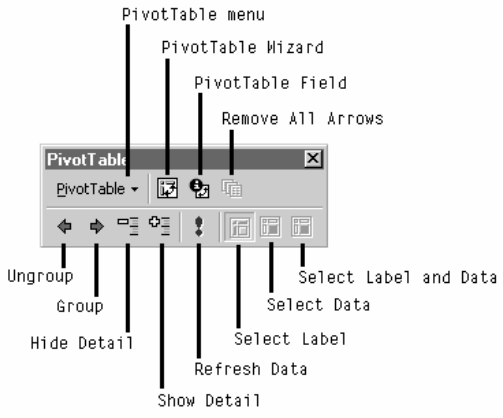
The balsakhi program had a sample of 123 schools in Vadodara. The schools had differences along the dimensions of language of instruction and gender of school according to the chart below:

Language of Instruction	Gender of School	Number of Schools
Gujarati	Male	18
Gujarati	Female	17
Gujarati	Both	70
Hindi	Both	10
Marathi	Both	8

The researchers anticipated that the balsakhi could have differential effects dependending on the language of instruction or the gender of the school; accordingly, they wanted to be sure that the treatment and comparison groups would have equal numbers across both of these characteristics. Say for example, the treatment group had a disproportionately low number of Hindi and Marathi schools. Then critics could have claimed that the improved test scores in the treatment schools reflected the absence of non-Gujarati schools and not the presence of the balsakhi.

Randomization without stratification can create unbalanced treatment and control groups. To see this, open the file stratify.xls and select the worksheet "Section I". Open the pivot table toolbar by going to "view" then "toolbars" and then selecting "pivot table" as shown below on the left.





The above toolbar should appear in Excel. When you select the chart "Randomized Allocation without Stratification," some of the toolbar icons should light up. This chart is generated from a list of the Vadodara schools and their characteristics. Schools have been randomly allocated to the treatment and control groups, and the chart shows this allocation by the characteristics gender and language.

When you select the "refresh data" option () from the pivot toolbar, this generates a new set of random numbers and re-allocates schools randomly to the treatment and control groups. Each time that you refresh you effectively run run a new randomization of the schools group assignment. Experiment with this a couple of times to prove to yourself that randomized allocation can lead to very unbalanced "randomly-allocated" groups in some circumstances.

Section II: Unstratified but Unbiased

Moving on to the worksheet "Section II", select the chart and again refresh the data. This chart shows the random allocation of schools in a fictional sample of 1000 Gujarati schools, with roughly one third of the schools male, one third female, and one third mixed gender type.

As you repeat the randomized allocation of the schools into control and treatment groups, note the breakdown of percent composition by the gender characteristic. How does this differ from Section I? Can you infer the reason for this change?

Section III: The Importance of Stratification

Moving on to "Section III," now select either of the two charts. These charts use the original (and actual) sample of Vadodara schools as in Section I. The graph on the left shows the results of the unstratified randomized allocation and the graph on the right stratifies the sample by language and gender prior to randomization.

As you refresh the data, note the difference in the two graphs. The stratified randomization is indeed updating. Different schools are being selected from the full sample into the treatment and control groups; however, this allocation occurs from separate "bins", in which schools share the same language and gender type within each "bin". By randomizing from pre-sorted groups, we are assured that the final treatment and comparison groups will be balanced across characteristics.

Section IV: Stratification – a Panacea?

Researchers also realized that the initial pupil teacher ratios could affect the way the program worked. These ratios ran from as low as 1 to as high as 45. If, on average, all treatment schools had much lower pupil teacher ratios prior to the program, critics could attribute the "balsakhi effect" to this fact – that the balsakhis were entering classrooms that were already favored due to their smaller class size.

Select "Section IV" and then either of the two charts. Randomize the group allocation a couple of times. As demonstrated by the chart on the right, the allocation of schools by their language

of instruction and gender categories will remain constant because we have stratified by these variables. How does the average pupil teacher ratio compare between the treatment and control groups by each gender-language category? How could you change this?

As noted in lecture, sometimes stratification is not possible. If you have several possible variables across which one could stratify, by attempting to stratify across all of them, the treatment and comparison groups may not be balanced across all of the groups. Researchers must select those characteristics, about which they are most concerned, that may influence the final outcomes in the treatment and control groups and stratify by these. Additionally, a variable by which researchers want to stratify may contain too many subgroups from which to randomize. Concretely, consider a case where the full sample is twenty and there are fifteen different languages in this group. Could you stratify in this case? If so, how would you implement this? If the variable was continuous (such as age and height) how might you randomize to get a balanced sample?

Section V: The Nitty-Gritty of Stratification with Excel.

Now that you understand the importance of stratification, try it out. There are three examples of varying difficulty in the file stratify.xls. The first "Data 1" is an abridged sample of thirty schools which needs to be stratified by language and then randomly allocated to treatment and control groups. The second "Data 2" is the full sample and requires stratification by language, gender, and an additional variable "portion of day". The final "Data 3" is the full sample and requires stratification again by three variables but the third is pupil teacher ratio now, a variable with relatively more possible values.

The following Excel tips may be of use:

USEFUL COMMANDS

Rand() -- *generates a random number*
Percentile(array,k) – *gives the kth percentile of the set of numbers in the array*
Rank(Number, ref) – *gives the ordinal rank of "number" in the set of numbers in the selection "ref"*
IF(logical test, value if true, value if false) – *returns particular values based on the results of the logical test*

More information about all the above commands is available through Excel help. The use of an array formula may also be valuable. See help under "array formula"

Also note that Excel has limitations with the sort command and can only sort three variables at a time. To sort by more variables, it is necessary to use visual basic to write a macro. The above exercises do not require the use of visual basic.

Stratification is also easily accomplished with statistical packages, such as STATA and SAS.

Poverty Action Lab
TRANSLATING RESEARCH INTO ACTION

Case 3: Women and Political Representation in India
Measuring the Effects of Reservations

This case study, with kind permission of the authors, is based on "Women as Policy Makers: Evidence from an India Wide Randomized Policy Experiment"; Raghabendra Chattopadhyay and Esther Duflo, Econometrica, 2004 as well as "Unappreciated Services: Performance and Perceptions"; Esther Duflo and Petia Topalova, MIMEO, MIT, 2004.

Throughout the world, women are underrepresented in government. In 2004, only 15.6 % of the parliaments of the world were composed of women. The low presence of women in government means that they are disempowered and have less influence over the decision making process.

To address this issue, political reservations for women have been implemented by many governments. It is not clear, however, what effect reservations have on the political process. Will candidates that come to power through reservations be less skilled? Will they be acting as proxies for other more powerful members of society? In systems where reservations exist for women, do policies, in fact, better represent the preferences of women? Do women and men really have different policy preferences? Legislation passed in India in 1992 provides a test case for addressing these questions.

Panchayats and Rural Self-Government

India is poor because the villages of India are poor. India will be rich if the villages of India are rich. Panchayats should be given greater power; for we want the villagers to have a greater measure of real swaraj [self-government] in their own villages. – Jawaharlal Nehru, first Prime Minister of India

Independence must begin at the bottom ... it follows, therefore, that every village has to be self-sustained and capable of managing its affairs. – Mahatma Gandhi

Panchayats are an old form of village government in India. When India became independent in 1947, Indian leaders wanted to empower them and incorporate them into the new country's system of governance. In fact an article was written in the constitution to give Panchayats the powers necessary to "function as units of self-government". However, in practice, political and financial support from the central government for the Panchayats disappeared in the 1950's and 1960's in most states. It was only in Karnataka, West Bengal, and Kerala, that Panchayats continued to be taken seriously.

The idea of Panchayat government was revived in the late 1980's when Rajiv Gandhi, then Prime Minister, began canvassing for rural support. A proposal to re-invigorate the Panchayats was widely popular, but it was defeated. Opposition political parties were strongly represented in the states and reluctant to give power to local village councils. In 1992, though, the bill was rewritten and passed. It became the 73rd Amendment to the Indian Constitution under Prime Minister Narasimha Rao. The rural development minister at the time observed "[this] casts a duty on the centre as well as the states to establish and nourish the village Panchayats so as to make them effective, self-governing institutions".

Seventy percent of India's population live in rural areas and are subject to the changes brought about by the 73rd amendment. The amendment sets guidelines for the states, which have the final authority to determine exactly what powers are delegated to the Panchayats. The 73rd amendment requires that the states supervise, every five years, direct elections for council members at the district, block and village levels.[1]

[11] In India, the district is the principal subdivision within states. There were 476 districts with an average population of 1.8 million in the early 1990s. The council at the district level is known as the Zilla Parishad. A block is a large subunit of a district. The block level council is known as the Panchayat Samiti.

The lowest level of the Panchayat is the Gram Panchyat (GP). The size of a GP varies from state to state but each GP covers between 1 and 12 villages, and between 2,000 and 10,000 people. Each ward, which may be a village, or a fraction of a village, elects a council member. A Pradhan (Panchayat Leader) is then elected, either directly or indirectly, by the council members. The councils must hold meetings, or Gram Sabhas, several times a year with a quorum of villagers in attendance. At one of these meetings, villagers must approve next year's budget and this year's activities.

The GPs are responsible for enacting schemes for economic development and social justice determined by the states in accordance with the 11th schedule of the act. The 29 areas of responsibility from the act are listed in the appendix.

The Panchayat is primarily responsible for implementing development programs, as well as identifying the needs of the villages under its jurisdiction. The major responsibilities of the GP are to administer local infrastructure such as public buildings, water and roads. The GP also identifies recipients for welfare programs such as widow's pensions, old age pensions and maternity pensions. The main source of financing is still the state, but most of the money which was previously earmarked for specific uses is now allocated through four broad programs.

(1) The Jawhar Rozgar Yojana (JRY) - infrastructure such as irrigation, drinking water, roads, repairs of community buildings, etc.
(2) A small additional drinking water scheme.
(3) Funds for welfare programs (widow's, old age and maternity pensions, etc.)
(4) A grant for GP functioning.

In many ways, the most striking part of the 73rd amendment is the institution of reservations for women. The framers of the amendment thought that the hierarchical nature of Indian village society would prevent all groups from having an equal voice (or a voice at all) in local governments. Women and scheduled castes and tribes would continue to be marginalized from the decision making process. In order to prevent this from happening, the amendment created a system of reserved seats for women and "backward" castes in the Panchayat council and council leadership positions.

The Debate Over Reservations

The idea of political reservations is a controversial one. First, there is the question of constitutional legitimacy. Since quotas restrict the election choices of voters, some consider them to be undemocratic. There is also the possibility that quotas reduce the quality of candidates running and result in less effective government. There is even a question about the premise on which quotas are based, that the identity of the elected representative matters for policy outcomes. It may be that, once elected, a candidate will support policies which reflect the preferences of the electorate or party rather than her preferences as a woman or member of a scheduled caste.

Those who favor quotas, however, feel they may make democratic institutions stronger. Some see unequal representation as the result of prejudices which themselves undermine the institutions of democracy. Exposing women and backward castes to decision making processes and representing them in positions of power is a way to lessen the prejudices that prevented these groups from being elected in the pre-quota system. Quotas may be necessary to make decentralization work for poor people. In a centralized administration, groups that are nationally disadvantaged can be organized in relatively powerful lobbying groups, and demand some specific redistributive measures. Many parties in Indian states represent the interest of the lower castes, for example. But at the village level, members of disadvantaged groups may be unable to fight for their rights. Decentralization may then lead to a worsening of their situation.

But this argument suggests that this situation would be affected by the identity of the Panchayat leader. It is not obvious that it should be the case. If the electoral and the participatory system is vibrant, those elected should reflect the policy preferences of the electorate, whatever their own identity. On the other hand, if it is weak, the nominal leader may be dominated by the local elite. This may be even more likely to happen if he or she happens to belong to a historically disadvantaged group.

The model in which candidates commit to policies reflecting the preferences of the electorate may not apply in practice. Strict campaign promises that are enforced are rarely if ever seen. Enforcement of campaign promises may occur through the mechanism of voting out incumbents or, in the Indian case, through the refusal to approve his budget at the community meetings. However, such enforcement is certainly not complete. In this case, one may expect that the identity of the policy maker will affect outcome.

Reservations for Women

"The transfer of power to one million women elected local representatives — many of whom are malnourished and illiterate — is the greatest social experiment of our time."

Joan Holmes, President, The Hunger Project

"You can tell the condition of a nation by looking at the status of its women."

Jawaharlal Nehru, first Prime Minister of India

The condition of women in India, relative to men, is by many measures very poor. According to the last census the ratio of women to men has declined to 93/100. Women work longer hours than men and have less access to education and health care. The maternal mortality rates in rural India are among the highest in the world. Similarly, malnutrition rates are high, and higher for women than men. According to an article published by UNICEF in the 1996 Progress of Nations "gender [in South Asia] has been the most statistically significant determinant of malnutrition among young children and a

frequent direct or underlying cause of death among girls below age 5". Women also earn less than men, and as in many other countries, their representation in government has been traditionally very low.

There is plenty of evidence from India and elsewhere that women have greater preferences for child health and nutrition than men. For example, the share of food in the budget tends to increase when the share of women in the family income increases. If this is the case, increasing women's participation in government is beneficial not only for reducing gender inequity but also to generate more favorable outcomes for children. Moreover, women's traditional responsibility to fetch water means that having women in parliament – thereby enhancing the decision making power of women – may lead to better water and sanitation infrastructure, a leading factor in reducing childhood mortality. As the president of the World Bank, James Wolfensohn, clearly puts it, empowering women is viewed by many as a "win-win" proposition:

"Education for girls has a catalytic effect on every dimension of development: lower child and maternal mortality rate; increased educational attainment by daughters and sons; higher productivity; and improved environmental management. Together, these can mean faster economic growth and equally important, wider distribution of the fruits of growth ... More education for girls will also enable more and more women to attain leadership positions at all level of society: from health clinics in the villages to parliaments in the capitals. This, in turn, will change the way societies will deal with problems and raise the quality of global decision making."

In addition, many believe that women would bring to government a cleaner style than men. Numerous experimental and survey-based studies find that women are more likely to exhibit "helping behavior" (Eagly and Crowley, 1986), take stronger stances on ethical behavior (Glover et al., 1997, Reiss and Mitra, 1998) and behave more generously when faced with economic decisions (Eckel and Grossman, 1998). The British Council conducted a study in 8 countries in East and Central Africa in late 2001, interviewing 800 people on their perception of women political leaders. The study found that more than 70 percent of people thought women politicians performed better than or equal to men, more than half of those interviewed felt that women politicians were less corrupt than men and cared more about basic community needs than men. Recent attention has also been drawn to the potential of advancing women as a means of fighting corruption. Swamy et. al. (2001) find that in the 43 countries surveyed in a ``world values survey," women are less likely to say illegal or dishonest behavior is justifiable. Swamy also finds that greater labor force participation and greater government involvement by women is related to less corruption.

These beliefs have translated into action in at least two other cases. In Mexico City, the police chief rescinded the ticket-writing authority of the city's 900 male traffic policemen and created a new force consisting exclusively of women, in an attempt to reduce corruption. A similar policy has been administered in Lima, Peru, where it is claimed there has been a fall in corruption after the introduction of women.

It was with these ideas in mind that the 73rd amendment has reserved seats in the Panchayats for women and scheduled castes and tribes. Specifically, the amendment

stipulates that one third of all council member seats are reserved for women. In addition, one third of all council leaders or Pradhan positions are reserved for women. Today, as a result of this amendment, there are about one million women elected to the village councils, or Gram Panchayats, in India.

The reservation of seats to women was determined on a random basis. In practice in most states, Panchayats reserved for women were selected randomly (every third GP in the order of their serial number), within each list of reservation status (reserved for Scheduled Castes, reserved for Scheduled Tribes, Non-reserved for SC or ST). Given this random allocation of seats all that is left for the researcher is to determine what data to collect, and how much of it.

While reservations have been part of the political landscape in India for a long time for scheduled castes and scheduled tribes since independence (there are reservations for political seats, government jobs, places in college, etc…), it is the first time that reservations were enacted for women. This experience was therefore closely watched, by proponents of reservations and their adversaries alike. Were women going to be effective leaders? Were they going to be less corrupt than men? Were other women going to be empowered by the presence of women politicians? Would the types of investments made by the Panchayat be different? Or would the new leaders simply act as proxies for their male relatives or other powerful people in the village, leading to less accountability and worst governance?

Field Observations

Some observers have questioned the possibility that women as leaders could have any effect, pointing out that many of those elected through the reservation system were illiterate and ignorant of the political process. It has also been argued that when the seats of male pradhans became reserved, they would simply nominate their wives who would continue to carry out their husband's policies. Finally, many of the newly-elected women faced strong prejudices due to their gender. In several particularly conservative states, acts of violence have been directed at women who undertook GP president responsibilities (Mathew and Nayak, 1996).

Yet, despite the opposition, some argue that:

"Elected women have proved to be more responsive, sympathetic and caring than their male colleagues. They are found to have greater integrity, more concern about issues like drinking water, health care and children's education in comparison to their male counterparts." (Mathew, 2001).

An evaluation by the NGO "Sahayi" of the performance of 6,200 women representatives in the various Panchayats in Kerala found that women elected in the second round of reservations were more educated, self-confident, and self-reliant than the previous batch.

Field researchers have suggested that the entry of women into the PRI (Panchayat Raj Instititutions) have had some effect on local government, and described reasons that may limit women leaders' effectiveness:

"In Madhya Pradesh women in the PRIs have made an impact in sectors like primary education, health, drinking water, watershed management, formation of self help groups, excise related issues like location of liquor shop etc. However, they have yet to make an impact in areas like agriculture, land improvement, minor irrigation, animal husbandry and rural industries, but this is more on account of the fact that the extension machinery is largely unable to address women or their needs"

Another field researcher noticed increased attention on women and girls' literacy, as well as reproductive health:

"One major outcome of women's entry into PRIs is a tremendous demand generation for female literacy and girl child education. The coming of women in the PRIs has helped focus attention on [issues of women's reproductive health] more sharply. In some Gram Panchayats with the help of NGO intervention, the women sarpanches have concentrated on the need for better health infrastructure in the rural areas"

Many also believe that the effect of the reservation policy will be different in different states. While it may be effective in states with a strong tradition of local democracy or in states where women are in a better position, many fear that, in the northern states where women are severely discriminated against, the policy is likely to be perverted and ineffective. To this, proponents argue that it is precisely in those states that the entry of women in politics may make the bigger difference.

The Project

Your evaluation team has been entrusted to determine the effects of reservations for women in the Panchayats. You want to go beyond the anecdotes, but at the same time, you do not want to limit your analysis to a simple count of what types of public goods are available in different Panchayats (although this definitely needs to be part of it). You would like the project to address all the dimensions in which reservations for women are changing the shape of local governance in India.

Discussion Topic 1:
What outcomes are likely to be affected by the reservation policy?

Before embarking on an evaluation exercise, you need to have a clear idea of what the likely impact of the program will be. It is on those dimensions that you think will be affected that you will try to collect data.

Given the background information, what are the main areas on which the reservation policy should be evaluated? In what different areas might we expect to see a difference as a result of reservations?

Box 1: Evaluation Area, Birbhum, West Bengal and Udaipur, Rajasthan

There are two districts of India where you can collect data, Birbhum in West Bengal and Udaipur in Rajasthan.

The district of Birbhum is located roughly 125 miles from the state capital, Calcutta. It is composed of two distinct regions. In the west is the mostly barren fringe of the Chota Nagpur plateau. In the east is the fertile, and densely populated, Gangetic plain. Birbhum had a population of 2.56 million in 1991 which increased to 3.01 million in 2001. Agriculture was the main economic activity; and rice, the main crop cultivated. The 1991 male and female literacy rates were 50% and 37%, respectively. In 2001 the female literacy rate rose to 60%. The district is known to have a relatively well functioning Panchayat system.

Unlike most states in India, West Bengal devolved powers to its Panchayats in 1978. Since then it has had a functioning Panchayat system with five year elections. Following the passage of the 1992 amendment, West Bengal gave additional responsibilities to its Gram Panchayats. Specifically, they were able to establish and administer informal education centers (SSK). The last Panchyat election took place in 2003. This was the second election where the reservation system was in place.

Udaipur district in Rajasthan is a predominantly "tribal" district, with a high poverty rate and a very low female literacy rate. The district has a population of 2.63 million according to the 2001 census. The population density of 196 persons per square kilometer is higher than the state average of 165 persons per square kilometer. However, more than 46% of the tribal population lives in remote villages accessible only by foot. Wheat and maize are the main cultivated crops and the district is rich in minerals like copper ore and limestone.

Early marriage of girls is prevalent in the district with nearly 70% marrying before turning 18. The sex ratio is 972/1000 reflecting the poor condition of women in Rajasthan in general. There are some indications that things are improving, however: between 1991 and 2001, the female literacy rate doubled from 20% to 44%.

In Rajasthan, the first Gram Panchayat elections were held in 1995 followed by a second election in 2000. Gram Panchayats in Rajasthan have the authority to spend money on local infrastructure but not to run their own schools.

There are several methods available to you and your team to answer questions about the effects of reservations for women in the Panchayats. Here are some examples of what could be done, with some ideas of the resources each of these activities would require, in term of man days (we will assume here that all other expenses can be computed using a simple overhead rule).

Pradhan Interview

Your team can interview the Pradhan. You can use this interview to collect information on a wide variety of subjects, from the Pradhan's background, education and political goals to the political involvement of her/his family members to the specific investments in public goods she has undertaken during her time in office. You can also ask to see the Pradhan's balance sheets, which are supposed to be public information.

A strength of interviewing the Pradhan and asking directly what investments he/she made in the GP is that you have a record for investments in all GP villages. You are not constrained to work with data from a subsample. Unfortunately, there is a serious issue of concern with this approach. The Pradhan's recollection of past policy initiatives may be imprecise and/or the Pradhan may lie. The cost of doing a pradhan interview is 1 man day/interview.

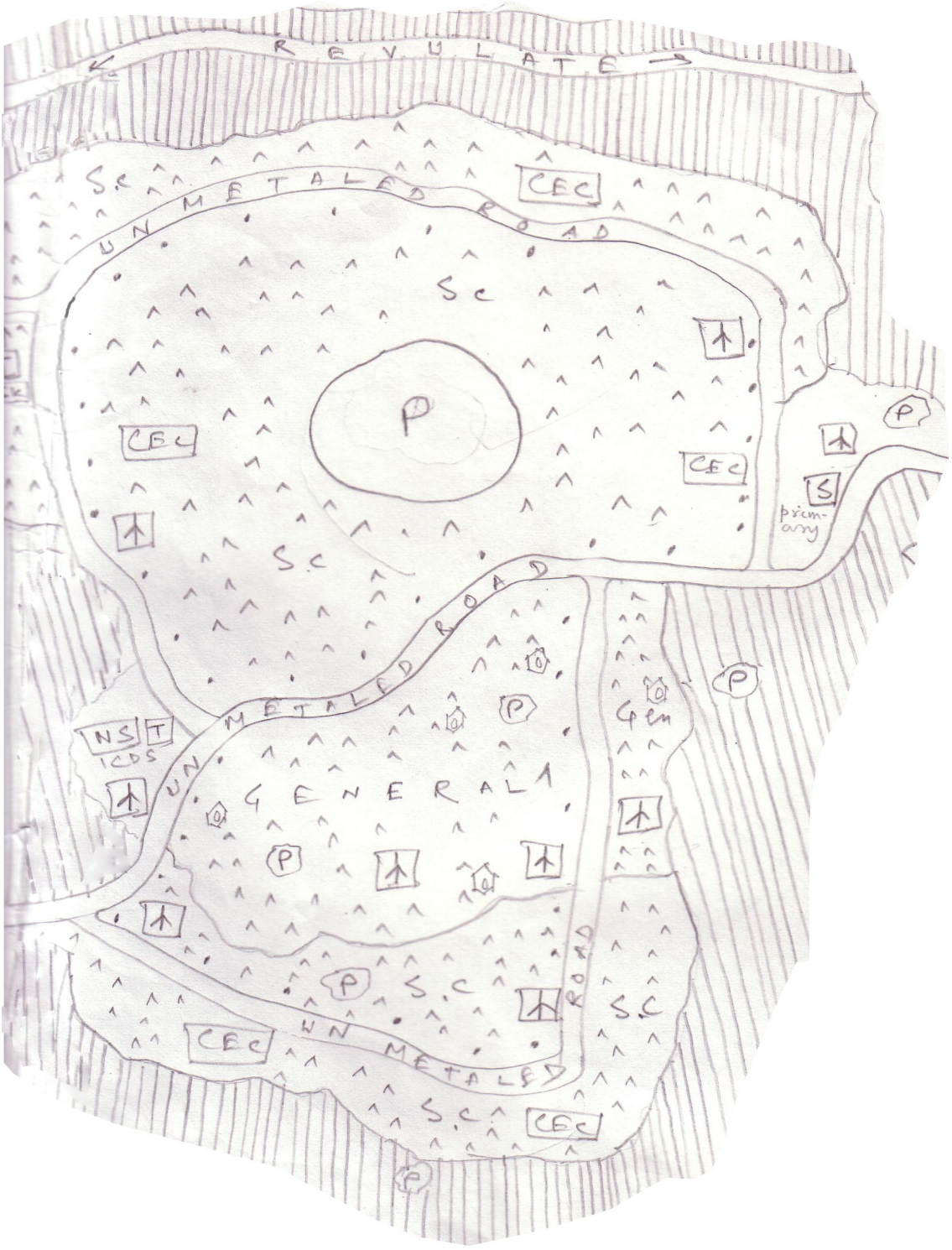
Participatory Resource Appraisal (PRA)

This is a method of gathering data at the village level. The PRA is a map of the village drawn with the help of 10 to 20 villagers. The map delineates schools, roads, wells, scheduled castes and tribe (SC and ST) areas, cultivated land and energy projects among other things. The villagers are asked to identify all available infrastructure in the village. It is also possible to ask, for each public good, when it was built or repaired. This is an extremely accurate method of gathering information on the public goods available in the village. An example of such a map is presented on the following page.

Moreover, the PRA is also an occasion to ask villagers general questions about the quality of different public goods in the village, about the participation of women and men in various activities (such as village meetings), and whether men or women from this village have complained about various issues to the Pradhan. At the PRA, you can obtain minutes of the past Gram Sabhas (this is much less detailed than the transcript), recording both attendance and the main issues raised. The drawback of such focus groups is that they do not represent the opinion of a representative sample of villagers.

Three man days are sufficient to complete a very detailed PRA and to complement it with focus groups. In addition, it will take one man-day to travel from one GP to the next, and half a man day to travel from one village to another within the same GP.

Participatory Resource Appraisal (PRA), Damdama Panchayat



Transcript of Gram Sabha

You could obtain more detailed information about what happened during the Gram Sabha by sending investigators to attend the meeting, and record what happened. They could both maintain an observation sheet (Who speaks? When? For how long?), as well as record the entire proceedings on a tape, which could be transcribed, translated, and that you can directly analyze. With this process, you will have a chance to observe in

detail the functioning of local democracy. An extract of such a recording (conducted in Karnataka) is shown in the appendix to this case. Attending the meeting and transcribing and translating takes a long time: you need to budget at least 5 man days for each Gram Sabha you decide to cover in this way.

Household Interviews

You also have the possibility to send a team of investigators to interview a sample of households in each village. Interviewing individuals would allow you to collect both objective information and perceptions from all households members.

Along with the PRA the household surveys allow very detailed data to be collected. However, this is the most expensive source of data. A short questionnaire (a simple questionnaire without physical measurement, and focusing on interviewing only one or two household members) will cost the research group roughly ½ man day per household, a long questionnaire (involving health measurement for example) will cost up to a full man day per household. Moreover, you need to start with a PRA to establish the household list from which your household can be sampled.

Box 2: Sample Size, Power and Clustering

When we select a sample of GP's for our study, we hope that the sample is representative of the general population. Otherwise, relationships that we find in our data may not hold in the 'true' population. If we randomly select GP's from the total GP population, then our sampling error will decrease as our sample size increases. So we may choose a sample size as large as our budget allows. Or we can work backwards and determine the sample size needed to gain a specific level of precision.

What affects the precision of our estimates besides the size of the sample? There is the variance of the characteristic we are trying to measure (for example, variance in investment in water) and the size of the treatment effect (how much does investment in water change with the treatment, i.e. with reservations for women). The more variance in the population, the larger the sample required to gain a given level of precision. Similarly, the smaller the treatment effect we are trying to detect, the larger the sample needed in order to detect it.

Power is a useful measure of how informative our sample is likely to be. It is defined as the probability with which we will notice a treatment effect when there is, in fact, a treatment effect in the sample.

Power varies with the method of randomization. If we randomize over GP's but are interested in measuring outcomes on individuals, households or villages within the GP, then the power of our sample decreases. This is because we expect there to be correlation between individuals, households or villages within a GP. When we use these 'sub-units' for analysis, we no longer have a truly random sample. We refer to randomization done on an aggregate of the unit of analysis as **clustered** randomization.

Choosing instrument and samples to answer policy-relevant questions

Your evaluation team may choose any number of villages or Panchayats from the districts of Birbhum and Udaipur and any combination of types of data collection listed above, provided the sampling method and the resource type do not exceed your budget.

Discussion Topic 2:
Choosing the right instrument to answer the right questions

Which of the possible data collection methods above can inform you of the impact of reservations for women on the dimensions (areas of interest) you identified in discussion topic 1? Which variables will you look at?

In determining what sampling method to use, your team will be concerned with the sample size required to get reasonable *power* for your analysis. The power is the probability that, if there is an effect of an expected size, you will be able to distinguish it from zero with a 95% confidence interval.

Discussion Topic 3: Power and Cost

Power Calculations

Start by completing the excel exercise.

How does power vary with sample size? With the effect size? With the way the sample is drawn? What is the tradeoff (in terms of budget) between choosing more villages within the same GP and choosing villages in different GPs?

Which effect on the number of investments in drinking water do you think your study needs to be able to detect to ascertain whether women Panchayat leaders did make a difference on the number of wells?

Given the magnitude of the effect you want to find, what is the smallest number of villages you need to detect the effect (from PRA)?

Does the study with the minimum number of villages necessarily have the smallest budget? Why or why not? How should you pick the villages to have the smallest budget to detect the effect? How many villages do you need?

Working within a budget

Given a budget of 900 man days, what data would you collect? Of the questions you are interested in which could you answer? Which questions that you would want to answer are you unable to answer within this budget?

Given a budget of 4000 man days, what data would you collect? What new questions could you answer?

Appendices

Appendix 1 - Responsibilities Listed in the 11th Schedule

Panchayat responsibilities. The 11 schedule of the 73rd amendment specifies 29 areas of responsibility that states may devolve to the panchayats. *Note: areas marked in **bold** are areas where independent observers have indicated that women bear primary responsibility.*

1. **Agriculture, including agricultural extension**
2. Land improvement, implementation of land reforms, land consolidation and soil conservation
3. **Minor irrigation, water management and watershed development**
4. **Animal husbandry, dairy and poultry**
5. **Fisheries**
6. **Social forestry and farm forestry**
7. **Minor forest produce**
8. Small scale industries, including food processing industries
9. Khadi (homespun cloth), village and cottage industries
10. Rural housing
11. **Drinking water**
12. **Fuel and fodder**
13. Roads, culverts, bridges, ferries, waterways and other means of communication
14. Rural electrification, including distribution of electricity
15. Non-conventional energy sources
16. **Poverty alleviation programmes**
17. **Education including primary and secondary schools**
18. Technical training and vocational education
19. **Adult and non-formal education**
20. Libraries
21. Cultural activities
22. Markets and fairs
23. **Health and sanitation, including hospitals, primary health centres and dispensaries**
24. **Family welfare**
25. **Women and child development**
26. **Social welfare, including welfare of the handicapped and mentally retarded**
27. Welfare of the weaker sections, and in particular, of the scheduled castes and the scheduled tribes
28. Public distribution system
29. **Maintenance of community assets**

The Hunger Project *http://www.thp.org March 14,2005*

Appendix 2: Plan Outlay for Annual Plan 2000- 01 for Schemes/Programmes of State Governments as per Eleventh Schedule

(Rs Crore)

Name of Scheme/Programme	Outlay
Crop husbandry	1268.77
Soil and water conservation	643.50
Animal husbandry	354.19
Dairy development	49.87
Fisheries	205.95
Forestry and wildlife	1393.03
Plantations	4.63
Agricultural research and education	292.21
Agricultural finance institutions	61.63
Other agricultural programmes	98.18
IRDP and allied programme	516.16
NREP/JRY	1161.22
Other programmes like EAS, etc	1062.46
Land reform	294.08
Minor irrigation	1907.93
Non conventional sources of energy	119.43
Village and small industries	800.87
Roads and bridges	8128.44
Road transport	1265.30
Inland water transport	16.74
Other transport services	614.50
General education	4972.07
Technical education	422.49
Art and culture	171.07
Medical and public health	3306.64
Water supply and sanitation	6031.74
Housing	1936.57
Welfare of SCs, STs and OBCs	1972.99
Social welfare	904.26
Nutrition	721.79
Other social services	119.86
Total	40818.65

Note: The outlay mentioned against each scheme/programme is the sum of the outlays for all the state governments allocated against each of their schemes.

Economic and Political Weekly January 10, 2004

Appendix 3: A Gram Sabha recording: Extract

Villager - Male (Speaker 1)
In our year 4 times the Panchayat and even 2 time Panchayat is also held. The sandy and market need is also described. But Panchayat needs do give all the support for the villagers for sandy and market.

[…]

G.P. Member - Male
We have put borewell but the pump is not put.

Villager - Male (Speaker 2)
Yes we will try and make it early for the borewell.
How many days you take for this.

G. P. Member - Male
We have taken good steps this but it is taking some time.

The sandy is also necessary but nobody is using this for selling or trading. So the place is not developed the shop are remaining like wise. We have done of MP quota but also we will not make that remain we will try and develop it as soon as possible

Villager - Male (Speaker 3)
Says, he is not treated properly as he is be looked low upon towards his caste.

G.P. Member - Male
But we are not saying so will you please be quite. We are not come to fight we are taking care.

Villager - Male (Speaker 4)
We have given application but what are you doing. We are very tired of giving our opinions we are very sick. We must say that government is not doing anything.

I think we must take some step against all these to government.

G. P. Member - Male
We have send the application to the HCF co-operation but we have not received any reply from the higher source so we have to wait.

Villager - Male (Speaker 4)
It is your duty to go and follow.

[….]

Villager - Male (Speaker 5)

How can you feel that happens, who is happy. There are nobody to ask you properly on the top of your voices, so you can do what ever you feel. You are taking all bribes you should give away the Panchayat you are taking 2000 Rs. Bribe you are taking 500 Rs and giving 1500 Rs to the higher ways.

How can you say that?

We have eye witness for Rs. 500 what you have taken not one we are many who have see you taking bribe.

Villager - Male (Speaker 6)
You are taking money and then saying you will do work.

G. P. Member - Male
How do you say.

Villager - Male (Speaker 7)
People who have given are here and people who have seen are here. We had given complaint to minister, but that was initially covered by taking bribe. They say that incident was not done at all. But we know that everything was done. Than he was transferred.

He is transferred.

Yes but why are you worried if he is transferred. What make you feel for that.

You should not tell all such this.

It is your faith its reality let everybody come to know. The bribe should not be entertained the transfer if such officer should be taken out first. Let the police come to the sabha when we do trouble I can face.

References

Chattopadhyay, Raghabendra and Esther Duflo. 2004. "Women as Policy Makers: Evidence from an India Wide Randomized Policy Experiment". Econometrica.

Constitution (Seventy-Third Amendment) Bill Insertion of New Part IX-A as Reported by Joint Committee. http://parliamentofindia.nic.in March 10, 2005.

Eagly, A.H., & Crowley, M. 1986. "Gender and Helping Behavior: a Meta-Analytic Review of the Social Psychological Literature". Psychological Bulletin, 100, 283-308.

Eckel, Catherine C. and Philip J. Grossman. 1998. "Are Women Less Selfish Than Men?: Evidence from Dictator Experiments." Economic Journal. 108, pp. 726-35.

Glover, Saundra H., Minnette A. Bumpus, John E. Logan, and James R. Ciesla. 1997. "Reexamining the Influence of Individual Values on Ethical Decision Making." Journal of Business Ethics, 16, pp. 1319-29.

Matthew, George. 2002. "Panchayatri Raj Institutions and Human Rights in India". International Council on Human Rights Policy.

Pal, Mahi. "Panchayati Raj and Rural Governance. Experiences of a Decade." Economic and Political Weekly. January 10, 2004. 137–143.

Ramalingaswami, Vulimiri, Urban Jonsson and Jon Rohde. 1996. "Commentary: The Asian Enigma". UNICEF, The Progress of Nations, Nutrition.

Reiss, Michelle C., and Kaushik Mitra. "The Effects of Individual Difference Factors on the Acceptability of Ethical and Unethical Workplace Behaviors." Journal of Business Ethics, 17(14), pp. 1581-93. (1998)

Swamy, Anand and Stephen Knack, Young Lee and Omar Azfar (2001). "Gender and Corruption," Journal of Development Economics, 64, 25-55.

Poverty Action Lab

TRANSLATING RESEARCH INTO ACTION

Case 4: Deworming in Kenya
Managing Threats to the Integrity of an Evaluation and Data Analysis

This case study, with kind permission of the authors, is based on "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities"; Edward Miguel and Michael Kremer, Econometrica Vol. 72, No 1 (January 2004) pp 159-217.

Introduction

Hookworm, roundworm, whipworm, and schistosomiasis infect one in four people worldwide. They are particularly prevalent among school-age children in developing countries. Intestinal worm infection rates are high throughout Kenya, especially among children of primary school age. A heavy worm load can lead to listlessness, diarrhea, abdominal pain, anemia, malnutrition, and even enlargement of the liver and spleen. In addition to the negative health and nutritional consequences, worm infections can result in impaired cognitive ability and poor academic performance. Low cost (50 cent) single dose oral therapies can kill most worms (although whipworm is sometimes harder to treat). Despite this, in many places worms are not treated on a regular basis, partly because they are seen as ubiquitous and non life-threatening and partly because reinfection is rapid and thus treatment must be regular. In the area of Kenya where the study discussed here took place, blood in stools (a symptom of worm infection) is considered a natural condition of childhood.

Scanning Electron Micrograph of Hookworm

Box 1. Disease Transmission: How Do Children Get Worms?

(Information from CDC, Division of Parasitic Diseases Hhttp://www.cdc.gov/ncidod/dpd/parasites/listing.htm#SH)

Geohelminths *(hookworm, roundworm, whipworm)*	Geohelminths are passed through contact with contaminated soil, generally by walking barefoot, or accidentally swallowing contaminated soil. Adult geohelminths live within the small intestine of their host and their eggs are passed in the feces, and often deposited into the soil.
Schistosomiasis	Schistosomiasis is caused by flukes, which have complex life cycles involving specific fresh-water snail species as intermediate hosts. Infected snails release large numbers of minute, free-swimming larvae (cercariae) that are capable of penetrating the unbroken skin of the human host. Even brief exposure to contaminated fresh water, such as wading, swimming, or bathing, can result in infection.

The Primary School Deworming Project

The Primary School Deworming Project (PSDP) was carried out by International Child Support (ICS) Africa (whose parent organization, Internationaal Christelijk Steunfonds, is a Dutch nongovernmental organization), in cooperation with the Busia District Ministry of Health office. The project took place in southern Busia, a poor and densely-settled farming region in western Kenya, in an area with the highest helminth infection rates in Busia district. The program involved 75 primary schools which were phased into deworming treatment in a randomized order. The 75 project schools consisted of nearly all rural primary schools in this area, and had a total enrollment of over 30,000 pupils between the ages of 6-18.

In January 1998, the 75 PSDP schools were randomly divided into three groups of 25 schools each: the schools were first stratified by administrative sub-unit (zone) and by their involvement in other non-governmental assistance programs, and were then listed alphabetically and every third school was assigned to a given project group. Due to ICS's administrative and financial constraints, the health intervention was phased in over several years. Group 1 schools received free deworming treatment in both 1998 and 1999, Group 2 schools in 1999, and Group 3 schools began receiving treatment in 2001. Thus in 1998, Group 1 schools were treatment schools, while Group 2 and Group 3 schools were comparison schools, and in 1999, Group 1 and Group 2 schools were treatment schools and Group 3 schools were comparison schools (Box 2). Treatment schools were treated regularly on the basis of standard protocols (Box 3).

Box 2. Treatment vs. Comparison Schools: Treatment Timeline

1998		1999		2001
Treatment	***Comparison***	***Treatment***	***Comparison***	***Treatment***
Group 1	**Group 2** **Group 3**	**Group 1** **Group 2**	**Group 3**	**Group 1** **Group 2** **Group 3**

Box 3. Deworming Drugs: Who Should Receive Them?

Given the minimal side effects of deworming drugs and the expense of individual screening, the World Health Organization (WHO 1992) has endorsed mass school based deworming programs in areas with high worm infection.

The WHO recommends that schools with geohelminth prevalence over 50 percent be mass treated with albendazole every six months, and schools with schistosomiasis prevalence over 30 percent be mass treated with praziquantel annually.

All treatment schools in this study met the geohelminth cut-off in both 1998 and 1999 and so were mass treated with albendazole. Six of 25 treatment schools met the schistosomiasis cut-off in 1998 and 16 of 50 treatment schools met the cut-off in 1999. Following standard practice at that time (Bundy and Guyatt 1996), the medical protocol did not call for treating girls thirteen years of age and older due to concerns about the potential teratogenicity (that is, threat of causing birth defects) of the drugs (WHO 1992). Since then, new evidence has suggested that this is not a threat and new WHO guidelines do not exclude girls of reproductive age.

Evaluation design: managing threats and analyzing data

In designing the evaluation, the authors faced a number of challenges related to i) the possible impact that treating children had on their school attendance, ii) the fact that not all children who were in the "treatment schools" got treated, and iii) the role children played in spreading worm infection. We will discuss each of these challenges separately.

Evaluation and attrition bias

When doing an evaluation of a program that targets school-age children and especially when such a program is carried out in schools, it is natural (and much more cost-effective) to collect data such as height, weight, and test scores on the school premises. Children are conveniently clustered together in one place at one time and data collectors avoid having to travel to the house of every child. However, if being infected by worms makes children listless and sick, it may affect children's enrollment in and attendance at school. This means that there may be more children attending treatment schools than comparison schools on the day that the evaluators show up to measure everyone's height, weight, and test scores (it should also be noted that pupil absence rates in Kenya are high to start with. This might have impacted their ability to measure the true impact of the program (see discussion topic 1).

Discussion Topic 1: Attrition Bias

Getting treated for worms meant that children turned up to school more often, and as a result, there were more children in treatment schools when the evaluators went to collect data at the end of the study.

Why could this potentially produce misleading results from the evaluation?

Are the children who benefited most from the deworming treatment different from the average child in the school? In other words, were the types of children who turned up in treatment schools, but who would have been absent if they had not been treated, just like an average child?

If these children were different from the average child, in what way are they likely to be different? Are they likely to be taller, smaller, more or less well-nourished, or to have higher or lower test scores than the rest of the children in their school?

Given this, would a simple comparison of outcomes for those who were at school when the evaluators arrived overestimate or underestimate the impact of the program?

How would you decide whether attrition bias might be a problem in another evaluation? What could the evaluators do to ensure they did not get misleading results?

Evaluation when not everyone who was meant to get treated actually gets the treatment?

Another complication was that not all the children in treatment schools actually received the treatment. One reason for this was that permission was required from parents and the community to treat the children. Individual permission from each child's parent was required because this was a research study. This would not have been necessary in Kenya if this had been a health program with no research component. Getting individual permission involved parents coming into the school and signing a permission slip in the Principal's office. This is not a trivial requirement for many households: traveling to school to sign the book may be time-consuming, and some parents may be reluctant to meet the headmaster when behind on school fees, a common problem in these schools. Another reason why not everyone got treated was that some children were absent on the day that the deworming medicine was given.

Discussion Topic 2: When the treatment group doesn't get treated

At first, it may appear natural that if someone who was randomly assigned to get treated doesn't for some reason get treated they should be excluded from the final evaluation. Indeed, this is what quite a number of randomized evaluations have done.

However, we have to ask the question "was it random that some of those in treatment schools did not get treated?" Are those who were meant to get treated, but didn't, likely to be different in any way from the rest of the children? Think about the reasons why children did not get treated. What are the characteristics of children for whom these factors are more important?

In what ways might this group be different from the average child?

If the evaluators had simply dropped those in the "treatment schools" who did not get the medicine from the study, would that have given them an accurate estimate of the impact of the program? Would they have overestimated or underestimated the impact of the program?

If the evaluators simply assumed that everyone who was meant to get treated got treated, what would this give them a measure of? What if only a very small percent of people who were meant to get treated actually got treated, would that cause problems for the evaluation in some circumstances? What if most of the people who were meant to get treated got treated, what would the resulting estimate be measuring? Are there circumstances when it would make sense to adjust this estimate for the number of people (in the treatment schools) who actually got treated? How might you go about doing such an adjustment?

It has also been the case in a number of studies that those who were assigned to the treatment group, but who were not treated, were grouped with the comparison group. Is this a better alternative?

Evaluation in the presence of externalities

Not only do school-aged children account for the bulk of infection, they are even more likely to spread the disease because they are less likely to use latrines and more generally have poor hygiene practices. This means that killing the worms in one child may help break the transmission of the disease and so improve the health of other children. This might happen relatively quickly because worms in water and in the bush can only survive for a relatively short period; in other words, the disease cycle is only maintained if new worms are regularly deposited.

Discussion Topic 3: Randomization with externalities within a group

Most of the previous work on evaluating the impact of deworming medication on children's health and school outcomes had ignored the fact that giving medication to one child may benefit another child by reducing their chance of infection.

Do you think that this previous work under- or overestimated the impact of deworming on children's health and education outcomes?

Solution to evaluation with externalities when the externalities are within a group

One way to solve the problem of estimating the impact of a program that involves externalities is to measure the effect at the community level rather than at the individual level. When talking about externalities, economists often use the example of a factory which produces smoke and is unpleasant for the whole community: it causes the washing of the neighbors to get dirty, and potentially causes ill health for many living around the factory. If you are trying to evaluate the benefit of reducing the smoke, you wouldn't just look at the costs and benefits to the factory itself, you would try to assess the benefits for everyone in the community.

In the same way, if there was one school in a village which everyone in the village went to and villages were far away from each other, it would be relatively easy to design an evaluation that took into account the fact that treating some children in the village might reduce the infection of others. You would simply randomize at the school/village level rather than the individual level and then measure the health benefits for all children in the village.

However, this is not the situation the evaluators faced in Busia District in Kenya. Instead of living in clustered villages, most people live on their farms which are spread relatively evenly, though one family is still close enough to and likely to infect/be infected by its neighbor. Many will have a choice of which school to attend and it is pretty frequent that children from neighboring houses go to different schools and even siblings in the same family go to different schools.

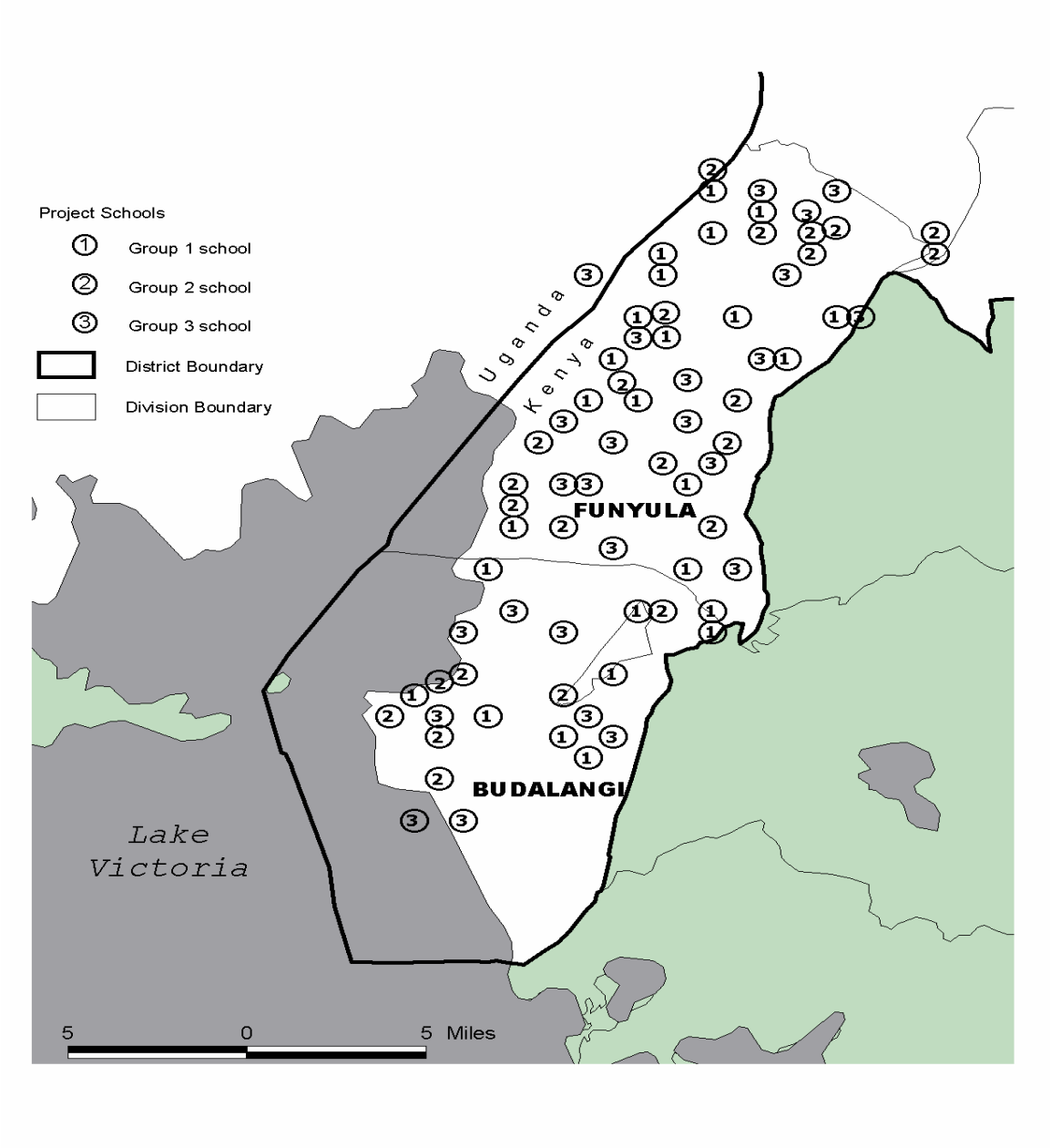
Discussion Topic 4: Randomization with externalities between groups

While the evaluators decided, because of the existence of externalities, to treat each school as a unit, there was still the issue that some children who were not in the "treatment school" might benefit from living near children who were treated.

How could the evaluators deal with this problem?

Look at the diagram on the next page which shows the geographical layout of all the schools. A "1" indicates that the school was in Group 1, a "2" that it was in Group 2, and so on. Does this give you any ideas of how to solve the problem and estimate the true impact of the program?

Primary School Deworming Project Schools, Busia District, Kenya



The GPS locations were collected before May 2000 (when the U.S. stopped intentionally downgrading GPS accuracy), and so may only be accurate to within several hundred meters, thus one school in this figure appears to be in Uganda. The school that appears to be in Lake Victoria is actually on a small island.

References

Bundy, D.A.P., and Guyatt, H.L. (1996). "Schools for Health: Focus on Health, Education, and the School-age Child." *Parasitology Today*, 12 (8), 1-16.

World Health Organization. (1992). *Model Describing Information. Drugs Used in Parasitic Diseases.* WHO, Geneva.