

Randomized Evaluations of Interventions in Social Service Delivery

By Esther Duflo, Rachel Glennerster, and Michael Kremer

What is the most effective way to increase girls' participation in school? How can we reduce the high absence rates of teachers in many developing countries? We know surprisingly little about these questions but finding answers is crucial to improving the quality of education in developing countries and the effectiveness of aid. Every year, millions of dollars are spent on evaluating development programs but these evaluations tend to focus on process: did the money go where it was meant to? How many teachers attended training courses? How many textbooks were delivered to schools? While tracking performance at this level is important, we should also be evaluating programs at a more fundamental level to find out whether, for example, training teachers or buying textbooks does more to raise test scores.

These fundamental questions, however, are hard to answer. Imagine, for example, that a new headmaster arrives at a school full of enthusiasm and new ideas. He wants to get the parents involved and sets up a parent committee. Word spreads that the new headmaster is good and some children transfer into the school from other local schools. If an attempt were made to evaluate the impact of the parents' committee it would be extremely difficult to disentangle the effects of the parents' committee, the impact of the headmasters enthusiasm on the other teachers in school, and the influx of new students who might be more motivated than average.

The cleanest and clearest way to establish how a program would work in an average school is to run a randomized trial. Choose 100 representative schools, initially establish a parents committee in half, phase in the program later in the other half, and in the mean time compare the outcomes in the two groups of schools. This approach, however, requires that evaluation is built into the design of the original program and that data are collected on all 100 schools which can be expensive. However, this is what we do if we want to know whether a drug or vaccine is effective and new research is showing this technique can teach us a lot about development.

Having strong evidence about what works is important for many reasons. Non-governmental organizations (NGOs) and governments can use this evidence to focus their limited budgets on those programs that are most effective. With widespread cynicism about the effectiveness of aid, providing clear evidence on the impact of different programs can also help galvanize support for more development assistance.

A series of recent studies have used randomized evaluations to address some of the fundamental questions raised at the beginning of this article. In this article, we summarize a few of the lessons that have already been learnt from successful randomized evaluations about how social service delivery and evaluation can be improved. More detail on these and other randomized evaluations can be found at www.povertyactionlab.com, the website of the new center at the Massachusetts Institute of Technology devoted to encouraging the use of randomized evaluation as a way to improve the effectiveness of

poverty programs in advanced and developing countries. While this article focuses on social service delivery, randomized evaluations can and have been applied in many other areas such as savings and credit, discrimination, and agricultural extension.

Improving access to social services: the example of school participation

The Millennium Development Goals call for universal primary school enrollment by 2015. However, until recently there were no good assessments of how best to achieve increased participation in education or how much it would cost.

Recent research suggests that a simple means of increasing school participation is by reducing the costs of schooling, or even paying for attendance. The PROGRESA program in Mexico provided cash grants to families if their children attended school regularly and received preventative health care like vaccinations. Schultz (forthcoming) takes advantage of fact that the program was phased-in in different areas randomly to assess its effectiveness. He finds an average increase in enrollment of 3.4 percent for all students in grades 1 through 8, and 14.8 percent among girls who had completed grade 6. In part because the randomized phase-in of the program meant the benefits were so clear, the Mexican government expanded the program, and similar programs are now being introduced elsewhere in Latin America.

Randomized evaluations performed in the same setting provide an opportunity to compare the cost effectiveness of various interventions. A series of evaluations in Kenya, provide a good example and provides comparative costs of different ways to increase

school attendance. The most cost-effective approach was to implement a twice-yearly school-based mass treatment with an inexpensive deworming drug. As the children's health improved so did their attendance and it cost only \$3.50 per additional year of school induced (Miguel and Kremer 2001). Providing school meals to pre-schoolers cost an average of \$36 per additional year of schooling induced, and also improved test scores in schools where teachers were well-trained prior to the program (Vermeersch 2002). In contrast, even under optimistic assumptions, a program which provided free school uniforms (among other inputs) cost \$99 per additional year of schooling induced (Kremer and others 2002).

Overall, these results suggest that school participation is quite elastic to cost and that school health programs may be one of the most cost-effective ways of increasing school participation. Is access to health care similarly elastic to cost? We will only know for sure when we have conducted a randomized evaluation of the policy.

Improving the quality of social services

While these results show that improving access to social services is relatively easy, other evaluations have shown that improving the quality of education is more difficult. Non-randomized evaluations suggest that providing additional inputs (such as textbooks or flipcharts) to under-resourced schools can improve learning and test scores (Glewwe and others 2002, Glewwe and others, forthcoming). However, results from a randomized evaluation point to a subtler picture: provision of textbooks increased test scores only among students who had scored in the top 40 percent on pre-tests prior to the program,

and did not affect scores for the bottom 60 percent of students. Flipcharts had no impact on test scores. This result shows how misleading the results of nonrandomized evaluations can be when, for example, inputs are more available in richer schools (as was the case with flipcharts in this study).

It would be wrong to conclude from these studies that providing inputs is necessary ineffective. Banerjee and others (2003) conducted a randomized evaluation of a remedial education program run by an Indian NGO. The program hired young local women from the communities to provide remedial education to students who were identified as lagging behind in traditional classes, and was found to have substantial positive impacts on learning, particularly for the weakest students. On average, student test scores increased by 0.39 standard deviations, and gains for the bottom third of students were 0.6 standard deviations, a very large impact. The program was shown to be at least 6 times more cost effective than computer assisted learning implemented (and evaluated) in the same schools.

These studies show that intuition about what works and what does not can be misleading. Substantial amount of money can be saved by finding out which inputs work and which does not.

Improving governance

One of the reasons why providing more inputs is not always effective is that social service delivery in developing countries is plagued by high absence rates and low

effectiveness. If teachers don't show up to school, the benefit of more textbooks may be limited to the few who can read well to start off with. How to reform the governance of social services, however, is a complicated and hotly debated question. Several recent studies involving randomized evaluations shed light on policies such as school reform, decentralization, and incentives.

A Colombian program provided vouchers for private schools through a random lottery (due to budgetary constraints), which thus allowed for credible estimates of program impact (Angrist and others 2002). Lottery winners were 15-20 percent more likely to attend private school, 10 percent more likely to complete 8th grade, and scored the equivalent of a full grade level higher on standardized tests (note that the vouchers were renewable conditional on satisfactory academic performance). The effects of the program were larger for girls than for boys. Winners were substantially more likely to graduate from high school and scored higher on high-school completion or college entrance exams. Overall, the benefits of this program to participants clearly exceeded the additional cost relative to the alternative of providing places in public schools.

In India, an important effort to decentralize the social good delivery started in the mid 1990. Local village councils, elected every five year, have the power to decide how to allocate expenditure on local public goods. The constitutional amendment that set up this program also required that one-third of all positions be reserved for women, and that a share equal to the representation of disadvantaged minorities (scheduled castes and scheduled tribes) be reserved for these minorities. To avoid any possible manipulation,

the reserved positions were randomly allocated allowing for a clear evaluation of the program. Chattopadhyay and Duflo (forthcoming) find that in villages reserved for women, the public goods chosen better reflect women's needs, and in villagers reserved for the scheduled castes a larger fraction of the public goods are allocated in the scheduled caste hamlet.

Conclusion

The results described above offer both substantive and methodological lessons. Inexpensive health programs, reducing the cost of school to households, or providing meals can substantially increase school participation. Given the existing governance problems in many developing countries, simply providing more resources may have a limited impact on social service quality, unless the resources are carefully allocated. Systemic reforms (such as school choice, or decentralization and involvement in decision making by disadvantaged groups) have important impacts.

These results also show that randomized evaluations of policy programs can be implemented successfully. They can take place in the context of small NGO or pilot programs, lotteries, or where a policy is phased in over time. The results can be in sharp contrast to conventional wisdom and the results of more traditional evaluations.

Finally, good evaluation promotes good policy. The positive results found through the rigorous evaluation of the PROGRESA program, led to the implementation of similar programs in other Latin American countries. The NGOs that have participated in the

evaluations discussed in this article have used the results to focus their resources where they are most effective.

Creating a culture in which rigorous randomized evaluations are promoted, encouraged, and financed has the potential to revolutionize social policy during the 21st Century, just as randomized trials revolutionized medicine during the 20th.

Esther Duflo is Professor of Economics, Department of Economics, MIT, and one of the founders of the Poverty Action Lab at MIT, eduflo@mit.edu.

Rachel Glennerster is Director of the Poverty Action Lab at MIT, rachel_glennerster@harvard.edu.

Michael Kremer is the Gates Professor of Developing Societies, Department of Economics, Harvard University, mkremer@fas.harvard.edu.

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer.

“Vouchers for Private Schooling in Colombia: Evidence from a Randomized Experiment.” *American Economic Review*, 92(5): 1535-1558, 2002.

Banerjee, Abhijit and Michael Kremer; with Jean Lanjouw and Peter Lanjouw. “Teacher-Student Ratios and School Performance in Udaipur, India: A Prospective Evaluation,” Mimeo, Harvard University, 2002.

Chattopadhyay, Raghavendra and Esther Duflo. “Women as Policy Makers: Evidence from an India-Wide Randomized Policy Experiment,” NBER Working Paper #8615, 2001, forthcoming in *Econometrica*, 2004.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Mimeo, Harvard University, November, 2002.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya," NBER Working Paper #8018, 2000, forthcoming in *Journal of Development Economics*.

Kremer, Michael, Sylvie Moulin, and Robert Namunyu. "Decentralization: A Cautionary Tale," Mimeo, Harvard University, March, 2003.

Miguel, Edward and Michael Kremer. "Worms: Identifying Impacts on Health and Education in the Presence of Treatment Externalities," *Econometrica*, 72(1): 159-217, 2004.

Shultz, T. Paul. "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program," Mimeo, Yale University, 2001, forthcoming in *Journal of Development Economics*.

Vermeersch, Christel. "School Meals, Educational Achievement, and School Competition: Evidence from a Randomized Experiment," Mimeo, Harvard University, 2003.