

American Economic Association

Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons

Author(s): Michael Kremer

Reviewed work(s):

Source: *The American Economic Review*, Vol. 93, No. 2, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003 (May, 2003), pp. 102-106

Published by: [American Economic Association](#)

Stable URL: <http://www.jstor.org/stable/3132208>

Accessed: 15/05/2012 04:30

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Economic Review*.

<http://www.jstor.org>

THE CONTRIBUTION OF RECENT INNOVATIONS IN DATA COLLECTION TO DEVELOPMENT ECONOMICS[†]

Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons

By MICHAEL KREMER*

This paper reviews recent randomized evaluations of educational programs in developing countries, including programs to increase school participation, to provide educational inputs, and to reform education. It then extracts some lessons for education policy and for the practice and political economy of randomized evaluations.

I. Increasing School Participation

Education is widely considered to be critical for development. The internationally agreed-upon Millennium Development Goals call for universal primary-school enrollment by 2015. However, until recently there were no good assessments of how best to achieve this goal or how much it would cost. Some argue that it will be difficult to attract additional children to school, since most children who are not in school are earning income that their families need. Others argue that children of primary-school age are not that productive, and modest incentives or improvements in school quality would be sufficient. Some see school fees as essential for ensuring accountability in schools and as a minor barrier to participation, while others argue that eliminating fees would greatly increase school participation.

The simplest way to increase school participation is to reduce the cost of school, or even pay for attendance. T. Paul Schultz (2003) takes

advantage of the randomized order of program phase-in to examine the impact of the Progreso program in Mexico, which provided cash grants to families conditional on their sending their children to school. He finds an average increase in enrollment of 3.4 percent for all students in grades 1 through 8; the increase was largest among girls who had completed grade 6, at 14.8 percent. In part because the randomized phase-in of the program allowed such clear documentation of the program's effects, the Mexican government decided to preserve and expand the program, and similar programs are being introduced elsewhere in Latin America.

One potential problem with conditional transfers is that the people administering the program may not enforce the conditionality in practice. School meals may therefore provide a stronger incentive to attend school than take-home rations ostensibly conditioned on school attendance (Amartya Sen, 2002). Christel Vermeersch and I are examining the effect of school meals on school participation in Kenya. School participation was 30-percent greater in 25 Kenyan preschools where a free breakfast was introduced than in 25 comparison schools. The provision of meals cut into instruction time, however. Overall, test scores were 0.4 standard deviations greater in the program schools, but only if the teacher was well-trained prior to the program (Christel Vermeersch and Kremer, 2002).

In many countries, parents face significant private costs of education, either for school fees or for other inputs, such as uniforms. In Kremer et al. (2002), Sylvie Moulin, Robert Namunyu, and I evaluate a program in which an NGO, International Christelijk Steunfonds Africa (ICS), provided uniforms, textbooks, and classroom construction to seven schools, randomly selected from a pool of 14 poorly performing candidate schools in Kenya. Parents are nor-

[†] *Discussants:* T. Paul Schultz, Yale University; Robert Townsend, University of Chicago; Duncan Thomas, University of California–Los Angeles.

* Department of Economics, Harvard University, Cambridge, MA 02138, Brookings Institution, and NBER (e-mail: mkremer@fas.harvard.edu). I am very grateful to Rachel Glennerster, Emily Oster, and International Christelijk Steunfonds Africa.

mally required to purchase uniforms at about \$6—a substantial expense in a country with per capita income of \$340. Dropout rates fell considerably in treatment schools, and after five years pupils in treatment schools had completed about 15 percent more schooling. In addition, many students from nearby schools transferred into program schools, raising class size by 50 percent. This suggests that students and parents were willing to trade off substantially bigger class sizes for the benefit of free uniforms, textbooks, and improved classrooms. The combination of these extra inputs and a 50-percent increase in class size led to no measurable impact on test scores. The cost savings from a much smaller increase in class size would have allowed the Kenyan government to pay for the uniforms, textbooks, and other inputs provided under the program, suggesting that existing budgets could be productively reallocated to decrease parental payments and substantially increase school participation.

Poor health may also limit school participation. To take one example, intestinal helminths (such as hookworm) affect a quarter of the world's population and are particularly prevalent among school-age children. Edward Miguel and Kremer (2003) evaluate a program of twice-yearly school-based mass treatment with inexpensive deworming drugs. Seventy-five schools were phased into the program in random order. Health and school participation improved not only at program schools, but also at nearby schools, due to reduced disease transmission. Absenteeism in treatment schools was 25 percent (or 7 percentage points) lower than in comparison schools. Including this spillover effect, the program increased schooling by 0.15 years per person treated. Gustavo Bobonis et al. (2002) find similar effects in India.

Because several programs were conducted in similar environments, cost-effectiveness can be readily compared. Deworming is extraordinarily cost effective, at only \$3.50 per additional year of school participation. In contrast, even under optimistic assumptions, provision of free uniforms would cost \$99 per additional year of school participation induced. The school feeding program, which targeted preschoolers rather than children of primary-school age, cost \$36 per additional year of schooling induced.

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.

II. School Inputs

Existing nonexperimental studies (e.g., Erik Hanushek, 1995) lead many to be skeptical about the effects of inputs on learning. These studies may be subject to omitted-variable bias, however. This bias could be upward, if observed inputs are correlated with unobserved parental or community support for education, or downward, for example, if compensatory programs provide assistance to poorly performing schools.

A randomized evaluation of a program in India that provided a second teacher, where possible female, to one-teacher schools, suggests that it did not improve test scores, although it did attract more girls to school (Abhijit Banerjee and Kremer, 2002). An ICS program that provided training and inputs to preschool teachers who started with minimal training also had little impact on test scores.

Although retrospective studies provide at best mixed evidence on the effect of many types of school inputs, they typically suggest that provision of additional textbooks in schools with low initial stocks can improve learning. Indeed cross-sectional and difference-in-difference analysis of Kenyan data would suggest that textbooks have dramatic effects on test scores. Results from a randomized evaluation paint a more subtle picture, however. Provision of textbooks increased test scores by about 0.2 standard deviations, but only among students who had scored in the top one or two quintiles on pretests prior to the program. Textbook provision did not affect scores for the bottom 60 percent of students (Paul Glewwe et al., 2002b).

Many students may have failed to benefit from textbooks because they had difficulty understanding them. Kenyan textbooks are in English, the official language of instruction, but English is most pupils' third language, after their mother tongue and Swahili. More generally, the Kenyan curriculum is set at a level that, while perhaps appropriate for elite families in Nairobi, is far ahead of the level typically attained by rural students, given high levels of student and teacher absence.

Given the results with textbooks, ICS tried providing an alternative input, flip charts, which

presumably were more accessible to weak pupils. Retrospective data from the area suggest that flip charts substantially improve test scores, but again, a randomized evaluation provides no evidence for this (Glewwe et al., 2003).

These examples suggest that the ordinary least-squares estimates are biased upward, rather than downward. This is plausible, since in a poor country with a substantial local role in education, inputs are likely to be correlated with favorable unobserved community characteristics.

III. School Reform

Given the limited measured impact of additional inputs on learning, focus has shifted to school reform initiatives, ranging from decentralization of budget authority, to strengthening links between teacher pay and performance, to vouchers and school choice. There is reason to believe that the Kenyan school system could benefit considerably from reform. Evidence discussed above suggests that budgets are misallocated and that the curriculum focuses excessively on the strongest students. Moreover, teacher incentives in Kenya, as in much of the developing world, are quite weak. Absence among teachers in our data is quite high, around 20 percent.

A decentralization program in Kenya that provided small grants to parent-run school committees induced them to purchase textbooks, with educational consequences similar to those of the textbook program. Providing larger grants led school committees to shift spending toward construction, and no educational impact could be observed from this, at least in the short run.

Some parent-run school committees in the area provide gifts to teachers whose students perform well. Glewwe et al. (2002a) evaluate a program that provided prizes to teachers in schools that performed well on exams and had low dropout rates. In theory, this type of incentive could lead teachers either to increase effort or, alternatively, to teach to the test. Empirically, teachers responded to the program not by increasing attendance, but by increasing prep sessions designed to prepare students for the exams. Consistent with a model in which teachers responded to the program primarily by increasing effort devoted to manipulating test scores, rather than by increasing effort at stim-

ulating long-term learning, test scores for pupils who had been part of the program initially increased but then fell back to levels similar to the comparison group at the end of the program.

Joshua Angrist et al. (2002) evaluate a Colombian program in which vouchers for private schools were allocated by lottery. Vouchers were renewable conditional on satisfactory academic performance. Lottery winners were 15–20 percent more likely to attend private school, 10 percent more likely to complete 8th grade, and scored 0.2 standard deviations higher on standardized tests (equivalent to a full grade level). 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/college-entrance exams. The benefits of this program to participants clearly exceeded the cost, which was similar to the cost of providing a public school place.

IV. Conclusion

The evaluations described above offer both substantive and methodological lessons. School participation can be increased substantially by inexpensive health programs, by reducing the cost of school to households, or by providing meals. Given the existing education system in Kenya, which like many developing countries has a curriculum focused on the strongest students, limited teacher incentives, and suboptimal budget allocation, simply providing more resources may have a limited impact on school quality. Decentralizing budgets to school committees or providing incentives based on test scores had little impact in Kenya, but a school-choice program in Colombia yielded dramatic benefits for participants.

These evaluations also provide methodological lessons:

- (i) As seen in the textbook and flip-chart examples, estimates from prospective randomized evaluations can often be quite different than the effects estimated in a retrospective framework, suggesting that omitted-variable bias is a serious concern.
- (ii) Many programs fail. Publication bias may therefore be substantial if only positive results are published. This may be particularly likely for retrospective evaluations.

Even with randomized evaluations, it is important to put institutions in place to ensure that negative results are disseminated.

- (iii) Randomized evaluations can shed light not only on the impact of specific programs, but also on behavioral parameters and questions of more general theoretical interest. For example, evaluations suggested that the Kenyan educational system is heavily geared toward top students and that reallocating budgets within primary education could lead to considerably better outcomes, pointing to perverse incentives created by Kenya's mix of local and national school finance.
- (iv) While randomized evaluations have their limitations, many of these either also apply to other methods or could be overcome with appropriate evaluation design. For example, provision of inputs might temporarily increase morale among students and teachers, and this could improve performance. While this would bias randomized evaluations, it would also bias fixed-effect or difference-in-difference estimates.

Another issue is that programs may create spillover effects on people who have not themselves been treated. These spillovers may be physical, as when deworming interferes with disease transmission, but they may also operate through prices, for example, when provision of school meals leads competing local schools to reduce school fees (Vermeersch and Kremer, 2002). If these spillovers are global, for example, due to changes in world prices, identification of total program impact will be problematic with any methodology. If spillovers are local, however, randomization at the level of groups can allow estimation of the total program effect within groups and can generate sufficient variation in local treatment density to measure spillovers across groups (Miguel and Kremer, 2003).

- (v) Randomized evaluations are feasible. They are labor-intensive and costly, but no more so than other data-collection activities. In the past, randomized evaluations have been rare events conducted by governments with multi-million dollar budgets. The evidence here suggests that another model is possible, with much cheaper eval-

uations of NGO projects. Unlike governments, NGOs are not expected to serve entire populations. Financial and administrative constraints often lead NGOs to phase in programs over time, and randomization will often be the fairest way of determining the order of phase-in. Even small NGOs can substantially affect budgets in developing countries. Given that many NGOs work in the developing world, and that they frequently seek out new projects, my experience has been that it is relatively straightforward to find NGOs willing to conduct randomized evaluations. Hitches are more often logistical than philosophical.

However, while NGOs are well-placed to conduct randomized evaluations, it is less reasonable to expect them to finance these evaluations. Given that accurate estimates of program effects are a public good, and in fact, largely an international public good, randomized evaluations should be financed internationally.

Conducting a series of evaluations in the same area allows substantial cost savings. Once staff are trained, they can work on multiple projects. Since data collection is the most costly element of these evaluations, cross-cutting the sample reduces costs dramatically. For example, the teacher incentives and textbook programs were evaluated in the same 100 schools: one group had textbooks only, one had textbooks and incentives, one had incentives only, and one had neither. The effect of the incentive program should thus be interpreted as the effect of an incentive program conditional on half the schools having extra textbooks. This tactic can be problematic, however, if there are significant interactions between programs, especially if one program makes the schools atypical.

I have argued that the problems of omitted-variable bias which randomized evaluations are designed to address are real and that randomized evaluations are feasible. They are no more costly than other types of surveys and are far cheaper than pursuing ineffective policies. So why then are they so rare? Lant Pritchett (2003) argues that program advocates systematically mislead swing voters into believing exaggerated estimates of program impacts. Advocates block

randomized evaluations since they would reveal programs' true impact to voters.

I would like to advance a complementary explanation, in which swing voters (or policy-makers) are not systematically fooled, but simply have difficulty gauging the quality of evidence. Suppose retrospective regressions yield estimates of program effects equal to the true effect, plus measurement error, plus a bias term, possibly with mean zero. Program advocates then select the highest estimates to present to policymakers, while any opponents select the most negative estimates. Knowing this, policymakers rationally discount these estimates. For example, if advocates present a study showing a 100-percent rate of return, the policymaker might assume that the true return is 10 percent. In this environment there is little incentive to conduct randomized evaluations. Since the resulting estimates include no bias term, they are unlikely to be high enough or low enough that advocates will present them to policymakers. Even if results are presented to policymakers, policymakers unable to gauge the quality of particular studies will discount them. Why fund a project that a randomized evaluation suggests has a 25-percent rate of return when advocates of competing projects claim a 100-percent rate of return?

Of course, if policymakers could distinguish the quality of studies, this problem would not arise. This suggests a potential role for a certification organization that would help policymakers identify credible studies.

Ideally, the World Bank and other development funders would require pilot programs and randomized evaluations before launching large-scale funding of new policies that are prone to evaluation, just as regulators require randomized trials before approving new drugs. In that case, randomized evaluations might play the same transformative role in social policy during the 21st century that they played in medicine during the 20th century.

REFERENCES

- Angrist, Joshua; Bettinger, Eric; Bloom, Erik; King, Elizabeth and Kremer, Michael. "Vouchers for Private Schooling in Columbia: Evidence from a Randomized Natural Experiment." *American Economic Review*, December 2002, 92(5), pp. 1535–58.
- Banerjee, Abhijit and Kremer, Michael. "Teacher-Student Ratios and School Performance in Udaipur, India: A Prospective Evaluation." Mimeo, Brookings Institution, Washington, DC, 2002.
- Bobonis, Gustavo; Miguel, Edward and Sharma, Charu. "Iron Supplementation and Early Childhood Development: A Randomized Evaluation in India." Mimeo, University of California-Berkeley, 2002.
- Glewwe, Paul; Ilias, Nauman and Kremer, Michael. "Teacher Incentives." Mimeo, Brookings Institution, Washington, DC, November 2002a.
- Glewwe, Paul; Kremer, Michael and Moulin, Sylvie. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Mimeo, Harvard University, November 2002b.
- Glewwe, Paul; Kremer, Michael; Moulin, Sylvie and Zitzewitz, Eric. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics*, 2003 (forthcoming).
- Hanushek, Eric A. "Interpreting Recent Research on Schooling in Developing Countries." World Bank Research Observer, August 1995, 10, pp. 227–46.
- Kremer, Michael; Moulin, Sylvie and Namunyu, Robert. "Unbalanced Decentralization." Mimeo, Brookings Institution, Washington, DC, November 2002.
- Miguel, Edward and Kremer, Michael. "Worms: Identifying Impacts on Education and Health in the Presence of Externalities." *Econometrica*, 2003 (forthcoming).
- Pritchett, Lant. "It Pays To Be Ignorant." *Journal of Policy Reform*, 2003 (forthcoming).
- Sen, Amartya. "Introduction," in *The Pratiche education report*. Pratiche India Trust, 2002.
- Shultz, T. Paul. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics*, 2003 (forthcoming).
- Vermeersch, Christel and Kremer, Michael. "Food for Education: Evidence from a Randomized Evaluation." Mimeo, Harvard University, November 2002.