The Role of Randomized Evaluations in Making Progress Towards Universal Basic and Secondary Education

Michael Kremer*

Although there has been tremendous progress in expanding school enrollments and increasing years of schooling in recent decades, 113 million children of primary-school age are still not enrolled in school (UNDP, 2003).[1]

This paper reviews what has been learned from randomized evaluations of educational programs about how best to increase school participation. I first outline the intuition behind the important role of randomized evaluations in obtaining credible estimates of the impact of educational interventions, review other non-experimental methods of evaluation, and present some evidence on the biases that can arise with such non-experimental methods. I then discuss two types of programs that have been found to be extremely cost-effective and that could be implemented even within a very limited budget: school-based health programs and remedial education programs (that take advantage of available inexpensive sources of labor). I then outline a series of programs aimed at lowering the costs of school, or even paying students for attending school, that could be implemented if a higher level of financial support is available and discuss the possibility of educational reform through school choice, which could be implemented given sufficient political will within a country. The paper concludes by drawing some

* Department of Economics, Harvard University; The Brookings Institution; Bureau for Research in Economic Analysis of Development (BREAD); Center for Global Development; and National Bureau of Economics Research (NBER). Contact: Littauer Center 207, Cambridge, Massachusetts 02138; mkremer@fas.harvard.edu. I wish to thank David Bloom, Joel Cohen, Donald Green, Laura Ruiz Perez, and two anonymous referees for helpful comments, and Heidi Williams for excellent research assistance.

general lessons about the contribution of randomized evaluations to understanding the cost-effectiveness of various interventions.

Given the widespread consensus on the importance of education and the several reviews of the impact of education on income and other outcomes, this paper focuses not on the effects of education but on issues internal to the education system. The scope of this paper is also limited in that it does not address interventions intended to improve the quality of education, such as computer-aided instruction, unless these interventions cut costs and thus free resources that can be used to expand education.

Why the Results of Randomized Evaluations Are Valuable

There are many difficulties inherent in evaluating the impact of educational programs, as discussed below. By avoiding many of the potential biases associated with other evaluation methods, randomized evaluations are able to offer clearer estimates of program impact.

The Program Evaluation Problem

Evaluations of educational interventions—if and when they do occur—are most often conducted as afterthoughts, and not as a planned part of the program. Thus, when an estimate of the impact of an educational intervention is available, the estimate is most often based on retrospective data that are generated by everyday (non-experimental) variation across schools and households. In retrospective studies, it is very difficult to address the essentially counterfactual questions: How would the individuals who

[1] For information on the "Education for All" initiative (which involves numerous organizations including UNESCO, UNICEF, the European Commission, and the World Bank), see UNESCO 2000, 2002.

participated in the program have fared in the absence of the program? How would those individuals who did not participate in the program have fared in the presence of the program?

The difficulties inherent in examining these questions are obvious. Consider a simple illustration: a program is implemented in Uganda that seeks to improve school enrollment rates by offering free school meals, the motivation being to create additional incentives for students to attend school, as well as to possibly impact other outcomes such as nutritional status. In theory, we would like to observe a given group of students in both the state of participating in the school meals program as well as the state of not participating, and keep all other factors (rainfall, economic shocks, *etc*.) equal. If the group of students could be observed in both states, the evaluation would be simple; we could compare the outcomes in each scenario and know exactly what the effects of the program were because all other factors were kept constant.

Given the impossibility of observing any group of students in both states, in practice we compare data on some set of outcomes (such as school enrollment rates) for program participants to data on the same set of outcomes for some similar group of individuals who were not exposed to the program. Obtaining credible estimates hinges critically on the establishment of this second group of individuals that is “similar” to the program participants. The idea is that this “comparison” group gives us an idea of what would have happened to the program participants (the “treatment” group) had they not been exposed to the program.

In practice it can be quite difficult to construct a credible comparison group retrospectively, because individuals who did not participate in a program are most often not a good comparison group for those who did; for example, participation may be voluntary or programs may be specifically placed in poor areas. Any differences between the treatment group and the comparison group can be attributed to two separate factors: pre-existing differences (the “bias” term) and the actual impact of the program. Because we have no reliable way to estimate the size of this bias, we typically cannot decompose the overall difference between the treatment and comparison groups into a treatment effect and a bias term.

Bias could potentially occur in either direction: for example, estimates may be biased upward if programs are placed in areas that are more politically influential, or biased downward if programs are placed in areas that have particular problems attracting children to school. Bertrand, Duflo, and Mullainathan (2004) provide evidence that even controlling for pre-existing levels of outcome variables may not be sufficient to address such concerns.

This problem of bias in program evaluations can be addressed by carefully planning the evaluation in advance in order to construct a credible comparison group. In particular, the bias disappears if the treatment and comparison groups are selected randomly from a potential population of participants (such as individuals, communities, schools, or classrooms). In randomized evaluations we can be assured that, on average, individuals who are exposed to the program are not different, by more than chance variation, from those who are not, and thus that a statistically significant difference between the groups in the outcomes affected by the program can be confidently attributed to the program.

Other Techniques to Control for Bias

By construction, randomized evaluations address the bias problem discussed above. In part because randomized evaluations are not always possible to conduct, researchers (most notably labor economists) have made significant progress in developing alternative techniques that control for bias as well as possible, such as regression-discontinuity design, difference-in-difference techniques, and propensity score matching (see Angrist and Krueger, 1999; Card, 1999; and Meyer, 1995). However, each of these non-experimental methods rests on assumptions that cannot be tested, and in practice these techniques may contain large and unknown biases as a result of specification errors. LaLonde (1986) finds that many of the econometric procedures and comparison groups used in program evaluations did not yield accurate or precise estimates; econometric estimates often differed significantly from experimental results. Although Heckman and Smith (1995) argue that there have been important developments in non-experimental evaluation methods since LaLonde's 1986 review, there is nonetheless strong evidence that in practice the results of randomized evaluations can be quite different from the estimates offered by other evaluation methods.

One strategy to control for bias is to attempt to find a control group that is as "comparable" as possible to the treatment group, at least along observable dimensions. This can be done by collecting as many covariates as possible and then adjusting the computed differences through a regression, or by "matching" the program and the comparison group through the formation of a comparison group that is as similar as possible to the program group. One such method, "propensity score matching," first predicts the probability that a given individual is in the comparison or the treatment group on the basis of all available observable characteristics, then forms a comparison group of people who have the same probability of being treated as those who were actually treated. The weakness of this method, as with regression controls, is that it hinges on the identification of all potentially relevant differences between treatment and control groups. In cases where the treatment is assigned on the basis of a variable that is not observed by the researcher (demand for the service, for example), this technique can lead to misleading inferences.

A second strategy is often called the “difference-in-difference” technique. When a sound argument can be made that, in absence of the program, trends in educational outcomes in regions receiving the program would not have differed from trends in regions not receiving the program, it is possible to compare the change in the variables of interest between program and non-program regions. However, this assumption cannot be tested. To ascertain its plausibility, one needs time series data from before the program was implemented to compare trends over a long period. One also needs to be sure that no other program was implemented at the same time, which is often not the case. Finally, when drawing inferences, one must take into account that regions are often affected by time-persistent shocks that may look like “program effects.” Bertrand, Duflo, and Mullainathan (2004) find that difference-in-difference estimations, as commonly performed, can severely bias standard errors: the researchers randomly generated placebo laws and found that with about twenty years of data, difference-in-difference estimates found an “effect” significant at the 5 percent level, for as many as 45 percent of the placebo laws.

In one example of where difference-in-difference estimates can be used, Duflo (2001) takes advantage of a rapid school-expansion program in Indonesia in the 1970s to estimate the impact of building schools on years of schooling attained and subsequent wages. Identification is possible because the allocation rule for the school is known (more schools were built in places with low initial enrollment rates), and because the cohorts participating in the program are easily identified (children twelve years or older when the program started did not participate in the program). The increased growth in years of schooling attained and wages across cohorts in regions that received more schools suggests that access to schools contributed to increased education. The trends, quite parallel before the program, shifted clearly for the first cohort exposed to the program, thus reinforcing confidence in the identification assumption. However, this identification strategy is not usually valid; often when the timing of a policy change is used to identify the effect of a particular policy, the policy change is itself endogenous to the outcomes it was meant to affect, thus making identification impossible (see Besley and Case, 2000).

Finally, a third strategy, called “regression-discontinuity design” (see Campbell, 1969) uses discontinuities that are generated by program rules in some cases to identify the effect of the program through a comparison of those who made it and those who “almost made it.” That is, if resources are allocated on the basis of a certain threshold, it is possible to compare those just above the threshold to those just below.

Angrist and Lavy (1999) use this technique to evaluate the impact of class size in Israel, where a second teacher is allocated every time the class size grows above 40. This policy generates discontinuities in class size when the enrollment in a grade grows from 40 to 41 (as class size changes from one class of 40 to one class each of size 20 and 21), 80 to 81, etc. Angrist and Lavy compare test scores in classes just above and just below this threshold, and find that those just above the threshold had significantly higher test scores than those just below. This difference can confidently be attributed to the class size because it is difficult to imagine that schools on both sides of the threshold have any other systematic differences. Discontinuities in program rules, when enforced, are thus sources of identification. However, such discontinuities are not often enforced strictly enough to generate discontinuities that can be used for identification purposes, especially in developing countries. For example, researchers attempted to use as a source of identification the discontinuity in a policy of the Grameen bank (the flagship microcredit organization in Bangladesh), which restricts lending to include only people who own less than one acre of land (Pitt and Khandker, 1998). In practice, Grameen bank lends to many people who own more than one acre of land, and there is no discontinuity in the probability for borrowing at the threshold (Morduch, 1998).

Identification problems with non-randomized evaluation methods must be tackled with extreme care because they are less transparent and more subject to divergence of opinion than are problems with randomized evaluations. Moreover, the differences between good and bad non-randomized evaluations are difficult to communicate, especially to policy makers, because of the many caveats that must accompany the results. These caveats may never be provided to policy makers, and even if they are provided they may be ignored. In either case, policy makers are likely to be radically misled. Although non-randomized evaluations will continue to be necessary, there should be a commitment to conduct randomized evaluations where possible.

Evidence That the Results of Randomized Evaluations May Differ From Other Estimates

Several studies from Kenya offer evidence that estimates from prospective randomized evaluations can substantially differ from estimated effects in a retrospective framework, which suggests that omitted variable bias is a serious concern. For example, a Kenyan study (Glewwe et al., 2004) examines the potential educational impacts of providing schools with flip charts, which are poster-sized charts with instructional material that can be mounted on walls or placed on easels.[2] This intervention covered 178 primary schools, half of which were randomly selected to receive flip charts on topics in science, mathematics, geography, and health. Despite a large sample size and two years of follow-up data from a randomized evaluation, the estimated impact of flip charts on students' test scores is very close to zero and completely statistically insignificant. This implies that the provision of flip charts had no effect on educational outcomes. In contrast, several conventional retrospective ordinary-least-squares (OLS) estimates, which presumably suffer from the omitted variable biases as discussed, show impacts as large as 0.2 standard deviations, or 5–10 times larger than the estimates based on randomized trials.

As discussed below, similar disparities between retrospective and prospective randomized estimates have been found in studies of the impact of de-worming in Kenya (Miguel and Kremer, 2003, 2004). These results are consistent with the findings of Glazerman, Levy, and Meyers (2002), who assess both prospective (experimental) and retrospective (non-experimental) methods in studies of welfare, job training, and employment-service programs in the United States, synthesizing the results of twelve design-replication studies. Their analysis finds that retrospective estimators often produce results dramatically different from the results of randomized evaluations and that the estimated bias is often large. They are unable to identify any strategy that could consistently remove bias and still answer a well-defined question.[3] I am not aware of any systematic review of similar studies in developing countries.

Effective Programs for a Limited Budget

In this section, I outline two categories of programs that have been found to be extremely cost-effective means of making progress towards universal basic education. First, I discuss evidence, gathered from randomized evaluations of school-based health programs in Kenya and India, that suggests that simple and inexpensive treatments for basic health problems such as anemia and intestinal worms can have dramatic impacts on increasing the quantity of schooling that students attain. Second, I discuss the results of the randomized evaluation of a remedial education program in India that has been found to be an extremely cost-effective means of delivering education, particularly for weak students.

It is worth briefly defining the terminology of "school participation" as used in this paper. In developing countries, many pupils attend school erratically and the distinction between a frequently absent pupil and a dropout is often unclear. Attendance rates can vary dramatically among individuals, and thus large differences in the quantity of schooling would be overlooked by considering only years of schooling. One attractive way to incorporate wide variation in attendance when measuring the quantity of

---

[2] For more information, see Glewwe et al. (2004), http://www.nber.org/papers/w8018.pdf.

[3] Two recent studies not included in the analysis of Glazerman, Levy, and Meyers (2002) are those of Buddlemeyer and Skoufias (2003) and Diaz and Handa (2003). Both studies use randomized evaluation results as a benchmark to examine the performance of non-experimental methods (regression-discontinuity design and propensity score matching, respectively) for evaluating the impact of the PROGRESA program, discussed below. They find that appropriate methods provide an accurate analysis in these cases, but it is not clear that appropriate methods could have been selected *ex ante*.

schooling is to focus on a more comprehensive measure of schooling, often called "participation." For any child, participation is defined as the proportion of days that he or she is present in school to the number of days that the school is open, over a given period (e.g. Miguel and Kremer, 2004). This can be applied over one or more years, or just for a few days for which reliable data are available. Participation differs from attendance because attendance is usually defined only for children officially enrolled in school, whereas participation includes all children in the appropriate age range.

School-Based Health Programs[4]

Poor health may limit school participation, especially in developing countries. Intestinal helminths (such as hookworm, roundworm, whipworm, and schistosomiasis) affect a quarter of the world's population, and are particularly prevalent among school-age children. Moderate-to-severe worm infections can also lead to iron deficiency anemia, protein energy malnutrition, and undernutrition. Below, I review evidence from the evaluations of school-based health programs in Kenya and India.

Available low-cost, single-dose oral therapies can reduce hookworm, roundworm, and schistosomiasis infections by 99 percent (Butterworth et al., 1991; Nokes et al., 1992; Bennett and Guyatt, 2000), and the World Health Organization (WHO) has endorsed mass school-based de-worming programs in areas with high helminth infections. Miguel and Kremer (2004) examine the impact of a twice-yearly primary school de-worming

[4] For more information on the studies discussed in this section, see Miguel and Kremer (2004), http://post.economics.harvard.edu/faculty/kremer/webpapers/Worms_Identifying_Impacts.pdf; Miguel and Kremer (2003), http://emlab.berkeley.edu/users/emiguel/miguel_networks.pdf; Kremer and Miguel (2003), http://dsl.nber.org/papers/w10324.pdf; and Bobonis, Miguel, and Sharma (2004), http://emlab.berkeley.edu/users/emiguel/miguel_anemia.pdf.

program in western Kenya, where the prevalence of intestinal worms among children is very high (with an estimated 92 percent of pupils having at least one helminth infection). The de-worming program was implemented by a Dutch non-governmental organization (NGO), Internationaal Chistelijk Steunfonds (ICS) Africa, in cooperation with a local District Ministry of Health office. Due to administrative and financial constraints, the health intervention was randomly phased in over several years.

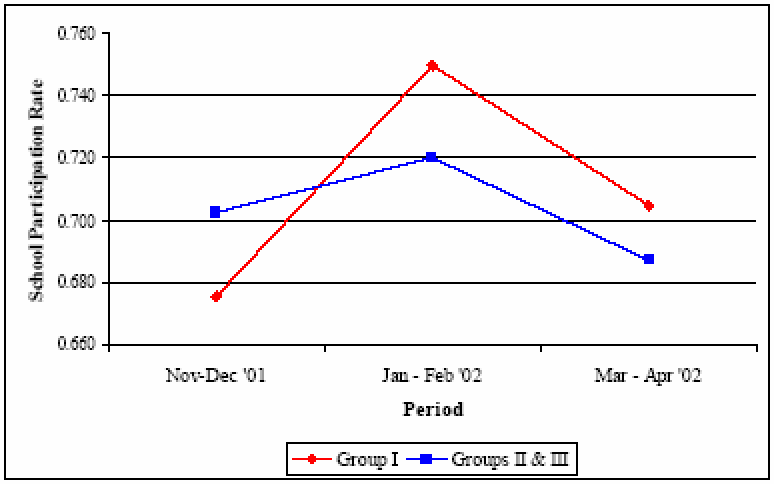
The authors find that child health and school participation improved not only for treated students but also for untreated students at treatment schools (measurable because 22 percent of pupils in treatment schools did not receive de-worming medicine) and untreated students at nearby non-treatment schools due to reduced disease transmission. Previous studies of the impact of de-worming fail to account for potential externalities, and Miguel and Kremer use two approaches to address identification issues that arise in the presence of these disease-reduction externalities. First, randomization at the level of schools allows them to estimate the overall effect of de-worming on a school even if there are treatment externalities among pupils within treatment schools. Second, cross-school externalities—the impact of de-worming for pupils in schools located near treatment schools—are identified using exogenous variation in the local density of treatment-school pupils generated by the school-level randomization.

Using this methodology, the authors find the direct effect of the de-worming program including within-school health externalities led to a 7.5 percent average gain in primary-school participation in treatment schools, a reduction in absenteeism of at least 25 percent. Including the cross-school externalities, the authors find that de-worming increased schooling by 0.15 years per pupil treated; decomposed into an effect of the treatment on the students treated and a spillover effect, school participation on average increased by 7.5 percent among pupils in treatment schools and by 2 percent among pupils in comparison schools. Including these externality benefits, the cost per additional year of school participation gained is only $3.50, making de-worming an extremely cost-effective method of increasing school participation.

Bobonis, Miguel, and Sharma (2003) also find evidence that school-based health programs can have substantial impacts on school participation. Iron deficiency anemia is another of the world's most widespread health problems, affecting approximately 40 percent of children in African and Asian settings (Hall et al., 2001). Bobonis et al. evaluate the impact of an NGO project in the slums of Delhi, India that delivers iron supplementation, de-worming medicine, and vitamin A supplements to 2–6 year old preschool students (an intervention that costs only $1.70 per child per year). Before the start of the project, 69 percent of children in the sample were anemic and 30 percent suffered from worm infections. Similar to the Kenyan de-worming program, the Delhi program was phased in randomly—in this case reaching 200 preschools over a period of two years. The authors found a sharp increase of 5.8 percent in preschool participation rates, a reduction in preschool absenteeism of roughly one-fifth. Effects were most pronounced for girls and children in areas of low socioeconomic status. The study also found large weight gains (roughly 1.1 pounds on average) within the first five months of the project. In combination with the findings from the Kenyan de-worming program, these results provide compelling evidence that school-based health programs can very cost-effectively increase school participation in low-income countries.

Figure 1. Iron Supplementation and De-worming Program in India: Pre-school Participation Rates Through Time. This table illustrates mean preschool participation

rates over time for students in the treatment and comparison groups, respectively. Group I received treatments from January–April 2002 and, as this graph illustrates, experienced a sharp increase in participation rates that remained greater than comparison rates through the end of year one.



Source: Bobonis, Miguel, and Sharma (2003).

These findings raise an important question: if school health programs can increase the quantity of schooling, how can such programs best be implemented in developing countries? Some contend that reliance on external financing of medicine is not sustainable and instead advocate health education, water and sanitation improvements, or financing the provision of medicine through local cost sharing. Kremer and Miguel (2003) analyze several de-worming interventions, including numerous "sustainable" approaches, such as cost sharing, health education, verbal commitments (a mobilization technique), and improvements in sanitation. They examine all except the sanitation efforts using randomized evaluations. Overall, their results suggest that there may be no alternative to continued subsidies for de-worming. The "sustainable" public health strategies of health education, community mobilization, and cost recovery were ineffective. For example, a small increase in the cost of the de-worming drugs led to an 80 percent reduction in take-up, relative to free treatment. On the other hand, provision of free de-worming drugs led to high drug take-up and large reductions in serious worm-infection rates. Miguel and Kremer find that the benefits of the health externalities alone are sufficient to justify not only fully subsidizing de-worming treatment, but also paying people to receive treatment.

In light of the preceding discussion of problems that arise with retrospective evaluation methods, I note that Miguel and Kremer (2004) find significant disparities between retrospective and prospective estimates of the de-worming project. For example, Miguel and Kremer estimate that students who were moderately or heavily infected in early 1999 had 2.8 percent lower school participation from May 1998 to March 1999. In contrast, an instrumental-variable specification (which imposes the condition that all gains in school participation result from changes in measured worm-infection status) suggests that each moderate-to-heavy infection leads to 20.3 percent lower school participation on average. The authors note several reasons why the instrumental-variable estimates are substantially larger, including issues with recurring infection, accounting for complementarities in school participation, and measurement error.

Remedial Education Programs[5]

Many developing countries have substantial numbers of educated, unemployed young people who could be cheaply hired to provide supplemental or alternative instruction in schools. Pratham, an Indian NGO, implemented a remedial education

[5] For more information, see Banerjee et al. (2003), http://www.povertyactionlab.com/papers/banerjee_cole_duflo_linden.pdf.

program in 1994 that now reaches over 161,000 children in twenty cities. Motivated by the belief that children often drop out of school because they fall behind and feel lost in class, the program hires young women from the communities (these women, the "Balsakhis," have the equivalent of a high school degree and are from the slum communities in which the schools are located) to provide remedial education in government schools. The Balsakhis teach children who have reached grade 2, 3, or 4 but have not mastered the basic grade 1 competencies. Children identified as lagging behind are pulled out of the regular classroom for two hours a day to receive this instruction.

Pratham wanted to evaluate the impact of this program, one of the NGO's flagship interventions, as they looked simultaneously to expand it. Expansion into a new city, Vadodara, provided an opportunity to conduct a randomized evaluation (Banerjee, Cole, Duflo and Linden, 2003). In the first year (1999–2000), the program expanded to forty-nine (randomly selected) of the 123 Vadodara government schools. In the following school year, the program expanded to all schools, but half received a remedial teacher for grade 3, and half received a teacher for grade 4. Grade 3 students in schools that received teachers for grade 4 served as the comparison group for grade 3 students who were directly exposed to the program. A similar intervention was conducted simultaneously in a district of Mumbai, where half the schools received the remedial teachers in grade 2, and half received teachers in grade 3. The program continued for an additional year, with each school switching the grade level to which the teacher was assigned. The program was thus conducted in several grades, in two cities, and with no school feeling that they were deprived of resources relative to others, because all schools participated in the program. After two years the program increased student test scores by 0.39 standard deviations, on average. Moreover, the gains were largest for children at the bottom of the distribution: those in the bottom third gained 0.6 standard deviations after two years. The impact of Pratham's program is increasing over time, and is very similar across cities and regardless of gender. The educational impact of the program, combined with data on the costs of teachers, suggests that hiring remedial education teachers from the community (at a cost of one or two dollars per child per year) appears to be twelve to sixteen times more cost-effective than hiring new teachers.

The positive effects of the program on children's academic achievement is remarkably stable across years and across cities, especially when the instability of the environment is considered—namely, that there was a major riot and catastrophic earthquake while the evaluation was being implemented. In their analysis, the authors carefully take into account these events, as well as their impacts on measures such as attrition, and treat that year of the program as a pilot program.

Table 1. Balsakhi Remedial Education Program in India: Estimated Cost Comparison of Balsakhis and Primary School Teachers in Mumbai. The costs of hiring Balsakhis is notably lower than the costs of hiring primary-school teachers, both in terms of cost in rupees per month and in terms of rupees per student.

		Rupees per month	**Rupees per student**
Balsakhi	Year 1	500	54
	Year 2	750	62
Primary school teachers	Years 1 & 2	7500	1318

Source: Banerjee, Cole, Duflo, and Linden (2003).

Promising Investments If Additional Resources Are Available

Several sources of evidence suggest that reducing the costs of education—or taking the further step of paying students to attend school—may significantly improve participation rates. In many developing countries, school fees and required inputs such as uniforms create significant private costs of education for parents. For example, in Kenya parents have historically been required to purchase uniforms that cost about $6, a substantial expense in a country with a per capita income of $340.

One might assume that a simple way to increase the quantity of schooling would be to reduce the cost of school or to pay students for school attendance. However, there is significant debate over the desirability of school fees. Proponents argue that fees are necessary to finance inputs, that they increase parental participation in school governance, and that the price elasticity of the demand for schooling is low (Jimenez and Lockheed, 1995). Opponents argue that school fees prevent many students from attending school and cite dramatic estimates from sub-Saharan Africa. When free schooling was introduced in Uganda in 1997, primary school enrollment reportedly doubled from 2.6 to 5.2 million children (Lokshin, Glinskaya, and Garcia, 2000); when primary school fees were eliminated in Tanzania in 2002, initial estimates were that 1.5 million students (primarily girls) reportedly began attending primary school almost immediately (Coalition for Health and Education Rights, 2002); and when Kenyan President Mwai Kibaki eliminated primary school fees in late 2002, the initial response was reportedly a massive influx of new students (Lacey, 2003). Although the elimination of school fees undoubtedly generated large increases in enrollments, the magnitude of the numbers cited in these journalistic accounts should be taken with a grain of salt for a number of reasons:

the underlying data on which they are based are often unclear; free schooling is sometimes announced simultaneous to other policy initiatives; and free schooling is often accompanied by a program that replaces school fees with per-pupil grants from the central government, which creates incentives for schools to overreport enrollments.

Evidence from several recent randomized evaluations suggests that programs designed to increase participation rates through a reduction in the costs of schooling, or even payments to students to attend school, can be effective. Below, I review evidence from the Mexican PROGRESA program as well as from a series of educational interventions in Kenya, including a school meals program, a program that provided school uniforms (among other inputs), and a girls' scholarship program.

Mexico's PROGRESA Program[6]

The PROGRESA program in Mexico distributed cash grants to women, conditional on their children's school attendance and participation in preventative health measures (nutrition supplementation, health care visits, and health education programs). When the program was launched in 1998, officials in the Mexican government took advantage of the fact that budgetary constraints limited their ability to reach the 50,000 potential participant communities of PROGRESA immediately. They instead started with 506 communities, half of which were randomly selected to receive the program while baseline and subsequent data were collected in the remaining communities (Gertler and Boyce, 2003). Another reason for this system of implementation was that it increased the probability of the program's continuation through shifts in political power, as proponents

[6] For more information on the PROGRESA program, see http://www.ifpri.org/themes/progresa.htm.

of PROGRESA understood that it would require continuous political support to be scaled up successfully.

The task of evaluating the program was given to academic researchers through the International Food Policy Research Institute (IFPRI), who made the data accessible to numerous researchers. A number of papers have been written on PROGRESA's impact, most of which are accessible on the IFPRI web site. The evaluations show that the program was effective in improving both health and education; in a comparison of PROGRESA participants and non-participants, Gertler and Boyce (2003) find that children on average had a 23 percent reduction in the incidence of illness, a 1–4 percent increase in height, and an 18 percent reduction in anemia. Adults experienced a reduction of 19 percent in the number of days lost due to illness. Schultz (2004) finds an average 3.4 percent increase in enrollment for all students in grades 1 through 8; the increase was largest among girls who had completed grade 6, at 14.8 percent.

School Meals Programs

In some contexts, the success of conditional transfers such as those awarded through the PROGRESA program may be undermined if the people administering the program do not enforce the conditionality in practice (Sen, 2002). In these circumstances, school meals may provide a stronger incentive to attend school because children must come to school to participate.

Government-subsidized school meals have been provided in India, Bangladesh, Brazil, Swaziland, and Jamaica in order to increase both enrollment and attendance (World Food Programme, 2002). Proponents of school meals also claim that school meals can increase both the quantity of schooling and academic performance by improving child nutrition. Critics argue that families may reduce resource allocation to children who receive school meals; however, if this were the case, school meals would still serve as an incentive for families to send children to school. Moreover, a retrospective study (Jacoby, 2002) from the Philippines suggests that parents do not reduce food provided at home in response to school feeding programs (see also Long, 1991, and Powell et al., 1983).

Vermeersch and Kremer conducted a randomized evaluation of the impact of school meals on participation in Kenyan preschools, and Vermeersch (2003)[7] finds that school participation was 30 percent greater in the 25 Kenyan preschools where a free breakfast was introduced than in the 25 comparison schools. There was some evidence that the provision of meals cut into instruction time. In schools where the teacher was relatively well trained prior to the program, the meals program led to higher test scores (0.4 of a standard deviation) on academic tests. There were no effects on tests of general cognitive skills, which implies that the school meals program did not improve children's nutritional status and that the academic test-score increases were likely due to the increase in time spent in school.

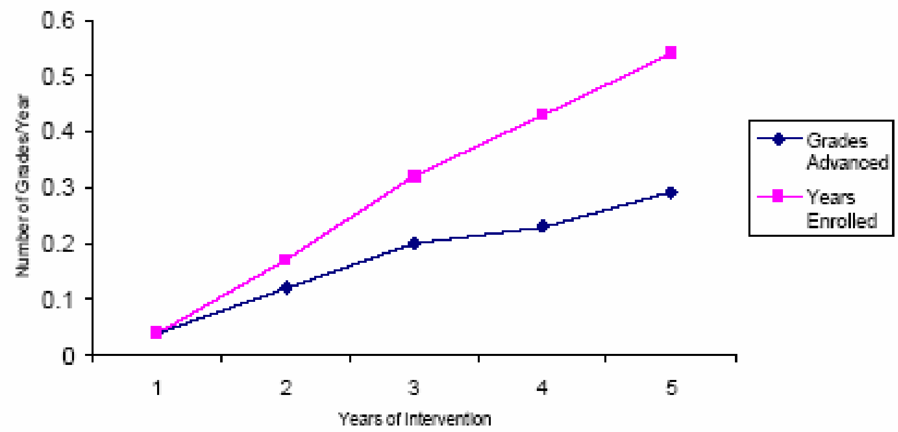
Provision of School Uniforms[8]

Kremer et al. (2002) conducted a randomized evaluation of a program in rural Kenya in which an NGO, Internationaal Christelijk Steunfonds (ICS) Africa, provided uniforms, textbooks, and classroom construction to seven schools that were randomly

[7] For more information, see Vermeersch (2003), http://www.nuff.ox.ac.uk/users/vermeersch/schoolmeals.pdf.
[8] For more information, see Kremer et al. (2003), http://post.economics.harvard.edu/faculty/kremer/webpapers/Decentralization.pdf.

selected from a pool of fourteen poorly performing schools. Dropout rates fell considerably in the seven schools that were randomly selected to participate in the program, and after five years pupils in those schools had completed about 15 percent more years of schooling. In addition, many students from nearby schools transferred into program schools, raising class size by 50 percent. This outcome suggests that students and parents were willing to trade substantially larger class sizes for the benefit of free uniforms, textbooks, and improved classrooms. The authors argue that the main reason for the increase in years of schooling was most likely the financial benefit of free uniforms. A separate randomized evaluation of a program which provided textbooks in Kenya (Glewwe et al., 2003) shows that textbooks had almost no impact on the quantity of schooling, and although the new classroom construction may have had an impact, the first new classrooms were not built until the second year of the program, whereas dropout rates fell dramatically in the first year. It is possible in theory that anticipation of later classroom construction affected participation, but the authors note that the presence of effects for students in the upper grades, who would have finished school by the time the classrooms were built, casts doubt on this argument.

Figure 2. Kenyan School Uniform, Textbook, and Classroom Construction Program: Program Effect on Grades Advanced and Years Enrolled. Given that the schools receiving the program were randomly selected, this graph illustrates the program effect by reporting the differences between the treatment and comparison groups over time.



Source: Kremer, Moulin, and Namunyu (2003)

Girls' Scholarship Programs[9]

In many countries there are significant gender disparities in access to education. It is estimated that about 56 percent of the 113 million school-age children not in school are girls, and in low-income countries there is a 9 percent gender gap in primary gross enrollment rates and a 13 percent gender gap at the secondary level (UNESCO, 2002). In sub-Saharan Africa, some studies estimate that barely 50 percent of girls complete primary school (Carceles, Fredriksen, and Watt, 2001). The question of how to increase enrollment rates of girls in primary and secondary schools in developing countries is often especially important.

There is some evidence that the elasticity of demand for schooling may be higher for girls than for boys, so policies and programs that do not specifically target girls may still result in greater increases in school participation for girls than for boys. Both Schultz (2004) and Morley and Coady (2003) find this trend in the evaluations of PROGRESA.

---

[9] For more information, see Kremer et al. (2004), http://emlab.berkeley.edu/users/emiguel/miguel_incentives.pdf.

The alternative is to implement programs that specifically target girls. Kremer, Miguel, and Thornton (2003) conducted a randomized evaluation of the Girls' Scholarship Program (GSP), which was introduced in rural Kenya in late 2001 to enhance girls' education. From a set of 128 schools, half were randomly chosen to be eligible for the program. The program consisted of a merit-based scholarship—one portion, intended for school fees, paid directly to the school and a second portion, intended for school supplies and uniforms, paid to the family—that rewarded girls in two districts of Western Kenya who scored in the top 15 percent on tests administered by the Kenyan government.

In the Busia district, the scholarship reduced student absenteeism by approximately 40 percent. Across all districts participating, the program increased the average probability of school attendance by 6 percent among girls in the first cohort of the program. It had a pre-program effect of 10 percent among girls in the second cohort in the year prior to their eligibility for the scholarships, possibly due to anticipation of the future scholarship opportunities or through peer effects. In addition, the test scores of girls eligible for the scholarship increased, by 0.2 standard deviations, as a result of the program. Moreover, schools offering the scholarship had significantly higher teacher attendance after the program was introduced, and scholarship winners were 7 percent more likely to rate themselves as a "good student" than girls who did not win scholarships.

Table 2. Girls Scholarship Program (GSP): Impact on School Attendance, Busia

	Dependent variable: Attendance in 2001, 2002 (boys and girls)
Program school	0.05** (0.02)

	Dependent variable: Attendance in 2001, 2002	
	Girls	Boys
Program impact		
Cohort 1 (2001)	0.06	0.08*
	(0.04)	(0.05)
Cohort 2 (2002)	0.01	-0.03
	(0.02)	(0.02)
Post-program impact		
Cohort 1 (2002)	0.02	0.02
	(0.02)	(0.03)
Pre-program impact		
Cohort 2 (2001)	0.10**	0.10*
	(0.05)	(0.06)
	Dependent variable: Teacher attendance in 2002	
Program school	0.05***	
	(0.02)	

Source: Kremer, Miguel, and Thornton (2003).

Notes: All estimates are ordinary least squares (OLS) estimates, marked as significantly different than zero at 90 percent (*), 95 percent (**), and 99 percent (***) confidence. Huber robust standard errors are in parentheses.

Other Educational Reforms: School Choice[10]

Given sufficient political will within a country, another possible educational reform aimed towards increasing enrollment is that of school choice. Angrist et al. (2002) examine the effects of Colombia’s voucher program on education outcomes. The program offered vouchers to attend private secondary schools to over 125,000 students from poor, urban neighborhoods. In most communities the demand for vouchers exceeded the supply, so voucher eligibility was determined by a lottery, generating a natural experiment. Data were collected from 1600 applicants for the vouchers (primarily from Bogota) three years after they had started high school. The sample was stratified so that half those sampled were lottery winners and half were lottery losers. Angrist and his co-authors find that lottery winners were 15–20 percent more likely to be in private schools, 10 percent more likely to complete grade 8, and that they scored 0.2 standard deviations higher on standardized tests than non-winners, equivalent to a full grade level.

A number of channels could account for the impact of the vouchers. First, lottery winners were more likely to have attended participating private schools, and these schools may be better than public schools. Second, vouchers allowed some pupils who would have attended private schools in the absence of vouchers to attend more expensive schools. Finally, voucher recipients who failed a grade risked losing their voucher, which increased the incentive to these students to devote more effort to school. The authors also find that vouchers affected noneducational outcomes: lottery winners worked less than lottery losers and were less likely to marry or cohabit as teenagers. Analysis of the economic returns to the additional schooling attained by winners after three years of participation suggests that the benefits likely greatly exceeded the $24 per winner additional cost to the government of supplying vouchers instead of public school places.

---

[10] For more information, see Angrist et al. (2002), http://papers.nber.org/papers/w8343.

Table 3. Colombia School Vouchers Program: Effects of the Bogota 1995 Voucher Lottery

	Coefficient on ever having used a private school scholarship (Bogota 1995 voucher lottery)		
Dependent variable	Non-lottery winner's mean	Ordinary least squares (OLS)	Two-stage least squares (2SLS)
Highest grade completed	7.5 (0.965)	0.167** (0.053)	0.196** (0.078)
In school	0.831 (.375)	0.021 (0.021)	0.010 (0.031)
Total repetitions since lottery	0.254 (0.508)	-0.077** (0.029)	-0.100** (0.042)
Finished 8th grade	0.632 (0.483)	0.114** (0.028)	0.151** (0.041)
Test scores (total points)	-0.099 (1.00)	0.379** (0.111)	0.291* (0.153)
Married or living with companion	0.016 (0.126)	-0.009 (0.006)	-0.013 (0.009)

Source: Angrist et al. (2001).

Notes: Results are from models which control for city, year of application, whether applicant had access to a phone, age, type of survey and instrument, strata of residence, and month of interview. Standard deviations are reported in parentheses for the non-lottery winner's means, and robust standard errors are reported in parentheses for the OLS and 2SLS columns.

Angrist, Bettinger, and Kremer (2003) suggest that the vouchers not only had significant effects on the short-run outcomes of recipients, but that their impact also persisted over time. Using administrative records of registration and test scores for a centralized college-entrance examination, the authors find that lottery winners were 7–8 percent more likely to take the university entrance exam (a good predictor of high school graduation given that 90 percent of all high school graduates take the exam), an increase of 15–20 percent in the probability of taking the exam. The authors also find an increase of 0.33 standard deviations in language test scores. Overall, these results point to a substantial gain in both high school graduation rates and achievement as a result of the voucher program. The size and persistence of these impacts suggest the voucher program was cost-effective.

One important concern about school vouchers is the effect of such programs on non-participants. For example, pupils left behind in public schools may be hurt by the departure of motivated classmates for private schools. Hseih and Urquiola (2002) use a retrospective fixed-effects estimation strategy to analyze Chile's nationwide school-choice program, and find that private enrollment rates negatively affect the relative test scores, repetition rates, and socioeconomic status of students in public schools. However, the author's retrospective identification strategy may be problematic, as private schools entered exactly where public schools were weak. On the other hand, voucher programs may enhance the education of non-participants if public schools may respond positively to increased competition (for evidence from retrospective studies, see Hoxby, 2000 and Bettinger, 2001). Such general equilibrium effects cannot be assessed by comparing lottery winners and non-winners, but both authors note that any negative external effects on non-participants would have to be extraordinarily large to outweigh program benefits.

Lessons

Several broad lessons can be drawn about the role of randomized evaluations in education policy, which I detail below. In addition, I briefly address some critiques of randomized evaluations that are frequently raised.

Costs

As is clear from the examples discussed in this paper, randomized evaluations are feasible and have been conducted successfully—they are labor intensive and costly, but no more so than other data-collection activities. The randomized evaluations discussed in this paper were conducted in concert with programs implemented by NGOs, and the cost-benefit estimates discussed include the costs to NGOs of program implementation.

Conducting evaluations in conjunction with NGOs has a variety of benefits. Once an evaluation staff is trained, they can work on multiple projects. Because data collection is the most costly element of these evaluations, crosscutting the sample can also dramatically reduce costs. For example, many of the programs seeking to increase school participation and learning were implemented in the same area, by the same organization. Of course, this approach must consider potential interactions between programs, which can be estimated if the sample is large enough, and may be inappropriate if one program makes the schools atypical. Another advantage of working with NGOs is that conducting a series of studies in the same area (such as the series recently conducted in Kenya) enhances comparability by allowing researchers to compare the cost-effectiveness estimates of different interventions in the same setting.

External Validity

Without a theory to explain why a program has the effect it has, it may be unwarranted to generalize from one well-executed randomized evaluation. However, similar issues of generalizability arise no matter what evaluation technique is used. One way to determine whether a program’s effects can be generalized is to encourage adapted replications of randomized evaluations in key domains of interest in several different settings. Although it will always be possible that a program unsuccessful in one context would have been successful in other adapted replications, replication of evaluations, if guided by a theory of why the program was effective, will go a long way toward alleviating concerns about generalizability.

The results of the first phase of a project often may be difficult to interpret because of circumstances that are unique to the first phase. If the project is unsuccessful, it may be because it faced implementation problems that could be avoided in later phases of the project; if the project is successful, it may be because more resources were allocated to it than would have been under a more realistic situation or in a less favorable context. Even if the choice of comparison and treatment groups ensures internal validity of estimates, the external validity of any method of evaluation may be problematic due to the specific circumstances of implementation—the results may not be able to be generalized to other contexts. Problems specific to randomized evaluations include the members of the treatment group changing their behavior (known as the Hawthorne effect) and members of the comparison group having their behavior affected (known as the John Henry effect) as a result of participation in the randomized evaluation.

Some of these concerns can be addressed by implementing adapted replications of successful (and potentially unsuccessful) programs in different contexts. Adapted replications present two advantages: first, in the process of "transplanting" a program circumstances change, and robust programs will show their effectiveness by surviving these changes; second, obtaining several estimates in different contexts will provide some guidance about whether the impacts of the program are notably different in different groups. Replication of the initial phase of a study in a new context does not necessarily entail a delay in the full-scale implementation of a program if the latter is justified on the basis of existing knowledge. More often than not, the introduction of a program must proceed in stages, and the evaluation only requires that participants be moved into the program in random order. Even within a single study, it is possible to check whether program effects vary with covariates; for example, a program may have differential effects in small and large schools.

One example of adapted replication is the work in India of Bobonis, Miguel, and Sharma (2003) who, as discussed previously, conducted an adapted replication of the de-worming study in Kenya. The baseline revealed that, although present, the levels of worm infection were substantially lower than in Kenya (in India, “only” 27 percent of children suffer from some form of worm infection). However, 70 percent of children had moderate-to-severe anemia; thus, the program was modified to include iron supplementation. The program was administered through a network of preschools in urban India. After one year of treatment, the researchers found a nearly 50 percent reduction in moderate-to-severe anemia, large weight gains, and a 7 percent reduction in absenteeism among 4–6 year olds (but not for younger children). This supports the conclusion of the de-worming research in Kenya (Miguel and Kremer, 2004) that school health programs may be one of the most cost-effective ways of increasing school participation and, importantly, suggests that this conclusion may be relevant in low-income countries outside of Africa.

A different external validity issue is that randomized evaluation may be unable to accurately predict the cost of a program if it were implemented on a broader scale. For example, if a program initially implemented by an NGO were scaled up, the relative increase or decrease in costs might be unclear due to issues of corruption, overhead, or supervision.

Issues That Can Affect Both Randomized and Retrospective Evaluations

Sample-selection bias, attrition bias, subgroup variability, and spillover effects can affect both randomized and retrospective evaluations. In the author's opinion, it is often easier to correct for these limitations when conducting randomized evaluations than when conducting retrospective studies.

Sample-selection problems could arise if factors other than random assignment influence program allocation. Even if randomized methods have been employed and the intended allocation of the program was random, the actual allocation may not be. For example, parents may attempt to move their children from a class or school without the program to one with the program. Conversely, individuals allocated to a treatment group may not receive the treatment (for example, because they decide not to take up the program). This problem can be addressed through intention-to-treat (ITT) methods or by using random assignment as an instrument of variables for actual assignment. The problem is much harder to address in retrospective studies because it is often difficult to find factors that plausibly affect exposure to the program that would not affect educational outcomes through other channels.

A second issue affecting both randomized and retrospective evaluations is differential attrition in the treatment and the comparison groups, where participants in the program may be less likely to move or otherwise drop out of the sample than non-participants. At minimum, randomized evaluations can use statistical techniques to bound the potential bias and can attempt to track down individuals who drop out of the sample (e.g. administer tests to students who have dropped out of school), which is often not possible with retrospective evaluations.

A third issue is subgroup variability, the possibility that a program will affect some individuals more than others. The issue of subgroup variability is important, but plausible theoretical mechanisms for its presence often exist. For example, Glewwe et al. (2003) find no evidence that provision of the official textbooks issued by the Kenyan government increased scores for the typical student. However, they do find evidence that textbooks led to higher test scores for the subset of students who scored well on a pretest. The authors note that English, the language both of instruction in Kenyan schools and of the textbooks, was the third language for most pupils. They cite evidence that many weaker pupils likely had difficulty reading the books.

Fourth, programs may create spillover effects on people who have not been treated. These spillovers may be physical, as found for the Kenyan de-worming program. De-worming interferes with disease transmission and thus makes children in treatment schools—and in schools near treatment schools—less likely to have worms, even if they were not themselves given the medicine. Spillovers may also operate through prices: Vermeersch (2003) finds that provision of meals in some schools led other schools to reduce school fees. Finally, there might also be learning and imitation effects (Duflo and Saez, 2003; Miguel and Kremer, 2003).

If spillovers are global (for example, due to changes in world prices), identification of total program impacts will be problematic with any methodology.

However, if spillovers are local, 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. For example, the solution in the case of the de-worming study was to choose the school (rather than the pupils within a school) as the unit of randomization, and to look at the number of treatment and comparison schools within neighborhoods. Of course, this requires a larger sample size.

One limitation of randomized evaluations is that the evaluation itself may cause the Hawthorne effect or John Henry effect. Although these effects are specific to randomized evaluations, similar effects can occur in other settings. For example, the provision of inputs could temporarily increase morale among students and teachers, which could improve performance. Although this would create problems for randomized evaluations, it would also create problems for fixed-effect or difference-in-difference estimates.

A final issue is that the program may generate behavioral responses that would not occur if the program were generalized. For example, children may switch into a school that is provided additional inputs. This may affect the original pupils by increasing class size, if class size affects the outcome of interest. Nationwide adoption of the policy would not have this effect.

Although randomized evaluation is not a bulletproof strategy, potential biases are well known and can often be corrected. This stands in contrast to most other types of studies, where the bias due to non-random treatment assignments could be either positive or negative, and cannot be estimated.

Conclusions and Avenues for Further Work

As illustrated by the substantive examples discussed above, a number of educational interventions have been shown to expand school participation quite effectively. Randomized evaluations of school-based health programs and remedial education programs suggest that these are extraordinarily cost-effective means of increasing the quantity of schooling attained in developing countries. Programs that reduce the cost of schooling or provide incentives for school attendance—whether implicitly, through school meals, or explicitly, through conditional grants—have been shown to have sizable impacts on school participation. Finally, school choice seems to have increased educational attainment in Colombia.

Randomized evaluations are needed on other means of increasing school participation rates, as there are a number of other promising avenues through which significant progress towards universal basic and secondary education can be made. For example, a recent study suggests there great potential likely exists on the margin of decreasing teacher absenteeism. A new representative survey of primary schools in India indicates that 25 percent of teachers in government primary schools are absent on a typical day. Two key interventions could take advantage of randomized evaluations: increasing community control in various ways (i.e., increasing the powers of parent-teacher associations) and increasing the frequency and quality of inspections, which preliminary evidence suggests can reduce teacher-absence rates.

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92 (5): 1535–1558.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2003. "Long-term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." forthcoming, *American Economic Review*.

Angrist, Joshua and Alan Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*, Vol. 3A, eds. Orley Ashenfelter and David Card. Amsterdam: North Holland.

Angrist, Joshua, and Victor Lavy (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement." *Quarterly Journal of Economics* 114 (2): 533–576.

Banerjee, Abhijit, Shawn Cole, Esther Duflo, and Leigh Linden. 2003. "Remedying Education: Evidence from Two Randomized Experiments in India." Mimeo. Massachusetts Institute of Technology.

Bennett, Andrew, and Helen Guyatt. 2000. "Reducing Intestinal Nematode Infection: Efficacy of Albendazole and Mebendazole." *Parasitology Today* 16 (2): 71–75.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Difference in Difference Estimates?" *Quarterly Journal of Economics* 119 (1): 249–276.

Besley, Tim, and Anne Case. 2000 "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110 (467): F672–F694.

Bettinger, Eric. 2001. "The Effect of Charter Schools on Charter Students and Public Schools." Mimeo. Case Western Reserve University.

Bobonis, Gustavo, Edward Miguel, and Charu Sharma. 2003. "Iron Deficiency Anemia and School Participation." Mimeo. University of California, Berkeley.

Buddlemeyer, Hielke, and Emmanuel Skofias. 2003. "An Evaluation on the Performance of Regression Discontinuity Design on PROGRESA." Institute for Study of Labor, Discussion Paper No. 827.

Butterworth, A.E., et al. 1991. "Comparison of Different Chemotherapy Strategies against *Schistosoma mansoni* in Kachakos District, Kenya: Effects on Human Infection and Morbidity" *Parasitology* 103: 339–344.

Campbell, Donald. 1969. "Reforms as Experiments." *American Psychologist* 24: 407–429.

Card, David. 1999. "The Causal Effect of Education on Earnings." Pp. 1801–1863 in *Handbook of Labor Economics,* Vol. 3A, eds. Orley Ashenfelter and David Card. Amsterdam: North Holland.

Carceles, Gabriel, Birger Fredriksen, and Patrick Watt. 2001. “Can Sub-Saharan Africa Reach the International Targets for Human Development?” Africa Region Human Development Working Paper Series. Washington, DC: The World Bank.

Coalition for Health and Education Rights. 2002. “User Fees: The Right to Education and Health Denied.” New York: CHER.

Diaz, Juan-Jose, and Sudhanshu Handa. 2003. “Estimating the Evaluation Bias of Matching Estimators Using Randomized-out Controls and Nonparticipants from PROGRESA.” Mimeo. University of North Carolina at Chapel Hill.

Duflo, Esther. 2001. “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment.” *American Economic Review* 91 (4): 795–814.

Duflo, Esther, and Michael Kremer. Forthcoming. “Use of Randomization in the Evaluation of Development Effectiveness.” Proceedings of the Conference on Evaluating Development Effectiveness, July 15–16, 2003. Washington, DC: World Bank Operations Evaluation Department (OED).

Duflo, Esther, and Emmanuel Saez. 2003. “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment.” *Quarterly Journal of Economics* 118 (3): 815–842.

Gertler, Paul, and Simone Boyce. 2003. “An Experiment in Incentive-based Welfare: The Impact of PROGRESA on Health in Mexico.” Royal Economic Society Annual Conference 2003, no. 85. Royal Econometric Society.

Glazerman, Steven, Dan Levy, and David Meyers. 2002. “Nonexperimental Versus Experimental Estimates of Earnings Impacts.” Mimeo. Mathematica Policy Research, Inc.

Glewwe, Paul, and Michael Kremer. Forthcoming. “Schools, Teachers, and Education Outcomes in Developing Countries.” In *Handbook on the Economics of Education*, ed. E. Hanushek and F. Welch, forthcoming.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2003. “Textbooks and Test Scores: Evidence from a Randomized Evaluation in Kenya.” Development Research Group, World Bank. Washington, DC: World Bank.

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. 2004. “Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya.” *Journal of Development Economics* 74: 251–268.

Hall, Andrew, et al. 2001. “Anemia in Schoolchildren in Eight countries in Africa and Asia.” *Public Health Nutrition* 4 (3): 749–756.

Heckman, James, and Jeffrey Smith. 1995. “Assessing the Case for Social Experiments.” *Journal of Economic Perspectives* 9 (2): 85–110.

Hoxby, Caroline. 2000. “Does Competition among Public Schools Benefit Students and Taxpayers?” *American Economic Review* 90 (5): 1209–1238.

Hseih, Chang-Tai, and Miguel Urquiola. 2002. “When Schools Compete, How Do They Compete?” Mimeo. Princeton University.

Jacoby, Hanan. 2002. “Is There an Intrahousehold Flypaper Effect? Evidence from a School Feeding Program” *Economic Journal* 112 (476): 196–221.

Jimenez, Emmanuel, and Marianne Lockheed. 1995. “Public and Private Secondary Education in Developing Countries.” World Bank Discussion Paper no. 309. Washington, DC: World Bank.

Kremer, Michael. 2003. “Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons.” *American Economic Review Papers and Proceedings* 93 (2): 102–115.

Kremer, Michael, and Edward Miguel. 2003. “The Illusion of Sustainability.” Mimeo. Harvard University.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2003. “Incentives to Learn.” Mimeo. University of California, Berkeley.

Kremer, Michael, Sylvie Moulin, and Robert Namunyu. 2002. “Decentralization: A Cautionary Tale.” Mimeo. Harvard University.

Lacey, Marc. 2003. “Primary Schools in Kenya, Fees Abolished, Are Filled to Overflowing.” *The New York Times*, January 7: A8.

LaLonde, Robert. 1986. “Evaluating the Econometric Evaluations of Training with Experimental Data.” *American Economic Review* 76 (4): 604–620.

Lokshin, Micahel, Elena Glinskaya, and Marito Garcia. 2000. “The Effect of Early Childhood Development Programs on Women’s Labor Force Participation and Older Children’s Schooling in Kenya.” Policy Research Report on Gender and Development, Working Paper Series no. 15. Washington, DC: World Bank.

Long, Sharon K. 1991. “Do the School Nutrition Programs Supplement Household Food Expenditures?” *The Journal of Human Resources* 26: 654–678.

Meyer, Bruce D. 1995 “Natural and Quasi-experiments in Economics” *Journal of Business and Economic Statistics* 13 (2): 151–161.

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

— — —. 2003. “Networks, Social Learning, and Technology Adoption: The Case of Deworming Drugs in Kenya.” Mimeo. Harvard University.

Morduch, Jonathan. 1998. “Does Microfinance Really Help the Poor? New Evidence from Flagship Programs in Bangladesh.” Mimeo. Princeton University.

Morley, Samuel, and David Coady. 2003. *From Social Assistance to Social Development: Education Subsidies in Developing Countries*. Washington, DC: Institute for International Economics.

Nokes, C., S. M. Grantham-McGregor, A. W. Sawyer, E. S. Cooper, B. A. Robinson, and D. A. P. Bundy. 1992. “Moderate-to-heavy Infection of Trichuris richiura affects Cognitive Function in Jamaican School Children.” *Parasitology* 104: 539–547.

Pitt, Mark, and Shahidur Khandker. 1998. “The Impact of Group-based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?” *Journal of Political Economy* 106 (5): 958–996.

Powell, Christine, Sally Grantham-McGregor, and M. Elston. 1983. “An Evaluation of Giving the Jamaican Government School Meal to a Class of Children.” *Human Nutrition: Clinical Nutrition* 37C: 381–388.

Schultz, T. Paul. 2004. “School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program.” *Journal of Development Economics* 74: 199–250.

Sen, Amartya. 2002. “The Pratichi Report.” Pratichi India Trust.

United Nations Development Programme (UNDP). 2003. *Human Development Report*. New York: UNDP.

United Nations Educational, Scientific, and Cultural Oraganization (UNESCO). 2000. *Informe Final, Foro Mundial Sobre la Educación*, Dakar, Senegal. Paris: UNESCO Publishing.

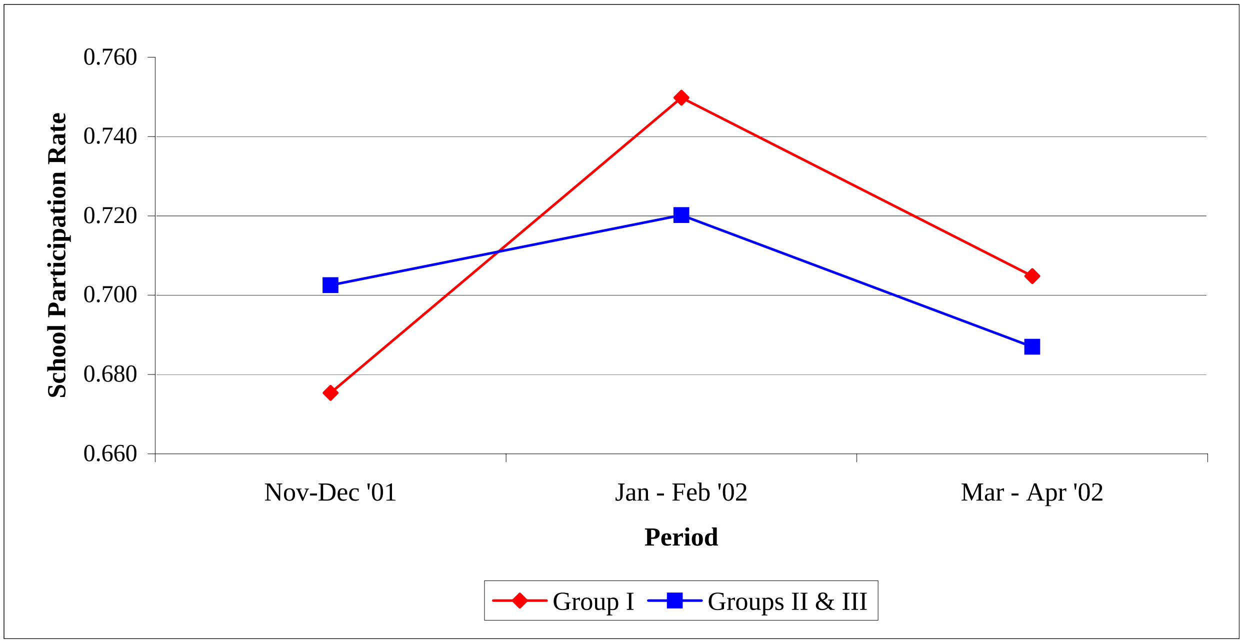
— — —. 2002. *Education for All: Is the World On Track?* Paris: UNESCO Publishing.

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

World Food Programme. 2002. “Global School Feeding Report 2002.” Rome: World Food Programme School Feeding Support Unit.

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

Figure 1: Pre-school Participation Rates through Time



Notes for Figure 1: Diamonds denote average pre-school participation for Group I, and squares are Groups II and III. The baseline (pre-treatment) period is November-December 2001. Group I was the treatment group during January-April 2002.

Table with Summary Statistics for Figure 1

Period	Group I	Groups II & III	Difference
Nov-Dec '01	0.675	0.702	-0.027
Jan - Feb '02	0.750	0.720	0.030
Mar - Apr '02	0.705	0.687	0.018