

## **Annex 1.7: Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya**

### **I. Introduction**

**Project Description.** Evaluating the effect of different types of education expenditure on student outcomes is particularly important in developing countries. Prior studies have suggested that the provision of textbooks is a cost-effective way of increasing test scores, and Kenya, with the extraordinarily scarce resources available to educators, makes a good case study. The evaluators note that only one in six children in grades 3, 4, and 5 has textbooks; this rises to one in four in later grades. In addition, physical facilities are extremely poor with many children sitting on the floor to learn.

The evaluation assessed the impact on learning outcomes of a 1996 program in which all grades in a randomly selected subset of 25 out of 100 rural Kenyan primary schools were provided with textbooks. English textbooks were given to grades 3 through 7, with a ratio of 6 textbooks to every 10 children; mathematics textbooks to grades 3, 5, and 7, with a 50 percent ratio; and science textbooks to grade 8, with a 60 percent ratio. In addition, each class was provided with a teacher's guide. Achievement tests were given to the students before textbooks were distributed and then again 10 months later. The same tests were also given to the control schools. This approach combines a randomized design with reflexive comparisons.

**Highlights of Evaluation.** This evaluation is an excellent illustration of developing and implementing a good survey design and then following that up with appropriate econometric techniques. It is particularly strong in showing how to draw inferences on level outcomes with stacked data, the use of difference-in-difference estimators, how to address selection and attrition bias, as well as measurement error and crowding-out issues. Another very interesting component of the evaluation is the focus on the intervention's impact on students in all parts of the distribution. Finally, the recognition and analysis of potential secondary effects is a very good example of looking at all dimensions of an intervention.

### **II. Research Questions and Evaluation Design**

The main focus of the research is to evaluate the effect of textbooks on learning outcomes. Because this is a complex concept, the outcomes are measured as the difference between textbook and comparison schools in

several dimensions: posttest scores, test score gains, differences between subject-grade combinations that did and did not receive textbooks, and child and teacher activity. The evaluation also considered other (often ignored) secondary effects, particularly the possibility that the provision of such a subsidy would reduce parental involvement, particularly in terms of crowding out other fundraising.

The evaluation design is quite complex. The Ministry of Education chose 100 needy schools for the intervention in 1995. These were divided into four groups—first on the basis of geography, then on an alphabetical basis within the geography. There was then an ordered assignment, on the basis of the alphabet, of each school to each of the four groups. Textbook assistance was staggered to go to the first group in 1996, the second group in 1997, and so on. Mathematics, English, and science textbooks were provided to different grades—primarily grades 3 through 7.

### III. Data

Math, English, and science exams were given to children in all these grades in each of the 100 schools before textbooks were distributed. The evaluation itself, however, makes use of pretests that were administered in grades 3 through 7 in October 1996 and posttests in October 1997. There are therefore data on some 8,800 students (in all grades) for each subject in the 100 schools and a total of over 26,000 observations. Because 25 schools received the textbooks in this period, students in these schools become the “textbook” group; the other 75 are the comparison group. In addition to test scores, data were also collected on school finances and on pedagogical methods.

Information on classroom utilization of textbooks was gathered by trained observers who visited each school and took minute-by-minute notes on eight possible classroom activities (ranging from general teacher and pupil activity to the use of textbooks by teachers and pupils). These notes covered 15 minutes and were then used to construct percentages of time spent by teachers and students in each different activity for a total of 551 class periods. Four to five students in each class were interviewed by field staff, who filled out a questionnaire on the basis of their responses.

Finally, data were gathered on school finances from a 1997 school and school committee questionnaire, which asked about fund-raising activities.

### IV. Econometric Techniques

It is worth noting the interesting issues generated by this sampling technique. Test scores within a school are likely to be correlated with each other, as are within-class scores. Similarly, test scores for different subjects

taken by the same child will be correlated. The intervention can also be evaluated in terms of the impact on outcomes on student learning levels or on student learning gains. In general, the effect of an intervention should be robust to different econometric techniques and different ways of looking at the data, and this was certainly the case here.

The evaluation proceeds by first providing estimates from a simple dummy-variable-level regression, with treatment dummies for each grade-subject combination with school, grade, and subject random effects (the dependent variable is the change in test scores from the pre- to the posttest). One attractive feature of this is that the dummies can be combined in very useful ways:

- Pooling several grades to estimate the impact of textbooks for a subject
- Pooling all test scores to estimate the average impact of textbooks for a grade; and
- Pooling all grades and subjects to estimate the weighted average impact of textbooks for all grades and subjects.

Clearly, the structure of the random effects varies with each approach, and the evaluation is very clear in this component.

The evaluation then proceeds with a difference-in-difference approach, which is relatively straightforward in that it simply compares post- and pretest scores between control and treatment schools.

The third approach, which is a little more complicated because it exploits within-school variation, deserves discussion. The regression applied here involves regressing test scores on dummies that capture whether the students were (a) in a textbook school and (b) in a subject-grade combination that received a textbook. This reduces problems introduced by school heterogeneity as well as sample selection problems—in the latter case because it captures the effect on test scores for the same student depending on whether or not the student received a textbook. It does assume, however, that test scores in different grade-subject combinations can be added and subtracted, and this very strong assumption may be the reason for very different results from this approach.

A recurring theme in evaluations is the desire to capture not just the average effect of the intervention but also the effect on subgroups of recipients. This evaluation provides a very useful illustration of the use of interaction terms and quantile regression. The former approach involves interaction between initial test scores and textbook dummies to capture the effect of textbooks on better versus poorer students, using both actual and instrumented values (initial test scores are correlated with the error term, causing a bias). The second approach, which involves using quantile regression, is also useful and increasingly popular. More specif-

ically, since least squares regression only captures the average impact of the textbook program, quintile regressions allow the effect of the treatment to vary depending on where the student is in the distribution.

The evaluation is also particularly strong in providing an application of how to look for selection and attrition bias. The major potential source of problems in this intervention is differential promotion and repetition rates between textbook and comparison schools. For example, children might be differentially promoted from grade 2 (a nontextbook grade) to grade 3 (a textbook grade) in textbook schools. Differential promotion biases down the results in the classes that the worst students are added to, and possibly biases up the results in the classes they came from. These two effects were captured in the evaluation by reestimating the model in two ways: dropping all repeaters from both sets of schools and dropping the worst students in each grade. The robustness of the results under both approaches confirmed the impact of the intervention.

Finally, in an illustration of considering the importance of secondary effects, the evaluation quantified the impact of textbook provision on parent fundraising. They found that the intervention did crowd out parent contributions—the amount of non-ICS aid received by comparison schools was \$465 and for textbook schools \$267 (the average value of ICS textbooks was \$485). They used simple regression analysis and also investigated, and confirmed, the hypothesis that smaller schools had more crowding out than larger schools.

**Who Carried It Out.** A Dutch nonprofit organization, International Christelijk Steunfonds, funded the project. The evaluation was performed by a Massachusetts Institute of Technology professor (Kremer) and two World Bank economists (Paul Glewwe and Sylvie Moulin). Some of the costs were covered by the National Science Foundation and the World Bank research committee.

## V. Results

The result of this evaluation was in marked contrast to the results of other evaluations of textbook interventions. The basic result was that there was no significant impact of textbooks on learning outcomes on average, but that there was a significant effect for better students. This was robust to different estimation techniques and cuts of the data.

## VI. Lessons Learned

The most useful lesson learned from this evaluation was the importance of using different econometric techniques to check for the robustness of

the empirical results. Even though the data collection was close to ideal, it is important that the estimated impact of the intervention remain roughly the same with different econometric assumptions and model specifications. The application of quantile regression and interaction terms was also a very useful way to analyze the impact on different subgroups of the population. Finally, it is important to look for and identify secondary effects—in this case, the potential for crowding out.

### **VIII. Source**

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 1998. "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Development Research Group (DECRG), World Bank, Washington, D.C. Processed.