CASE STUDY 2: FLIPCHARTS & CLASSROOM INPUTS IN KENYAN SCHOOLS

Why Randomize?

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

J-PAL thanks the authors for allowing us to use their paper.

KEY VOCABULARY

Counterfactual: what would have happened to the participants in a program had they not received the intervention. The counterfactual cannot be observed from the treatment group; can only be inferred from the comparison group.

Comparison Group: in an experimental design, a randomly assigned group from the same population that does not receive the intervention that is the subject of evaluation. Participants in the comparison group are used as a standard for comparison against the treated subjects in order to validate the results of the intervention.

Program Impact: estimated by measuring the difference in outcomes between comparison and treatment groups. The true impact of the program is the difference in outcomes between the treatment group and its counterfactual.

Baseline: data describing the characteristics of participants measured across both treatment and comparison groups prior to implementation of intervention.

Endline: data describing the characteristics of participants measured across both treatment and comparison groups after implementation of intervention.

Selection Bias: statistical bias between comparison and treatment groups in which individuals in one group are systematically different from those in the other. These can occur when the treatment and comparison groups are chosen in a non-random fashion so that they differ from each other by one or more factors that may affect the outcome of the study.

Omitted Variable Bias: statistical bias that occurs when certain variables/characteristics (often unobservable), which affect the measured outcome, are omitted from a regression analysis. Because they are not included as controls in the regression, one incorrectly attributes the measured impact solely to the program.

INTRODUCTION

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

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

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

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

SCHOOL QUALITY AND EDUCATIONAL RESOURCES

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

FROM TEXTBOOKS TO FLIPCHARTS

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

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

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

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

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

FIGURE 1: MATH FLIPCHART

OUTCOME MEASURE:

The process of evaluation requires more than just identifying inputs (in this case, flipcharts). Equally important is deciding what the intended impact is supposed to be and how to measure it. Interested in whether flipcharts improve competency level or learning, we need a measurement tool that quantifies this. Typically, educational outcomes are measured by standardized assessment tests. In evaluating educational inputs, the assumption is that any input that affects learning or knowledge should be reflected in test scores.

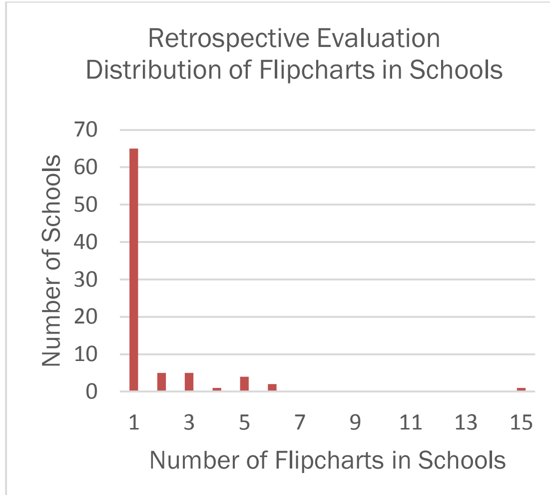
In Kenya, 8th grade is a pivotal year for students. At the end of the year students take an exam, the results of which determine whether they receive the Kenya Certificate of Primary Education (KCPE). Only with this certificate, can students proceed to secondary school. This exam is highly competitive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

OBSERVATIONAL STUDY

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

Of the 100 schools in the textbook evaluation, 83 schools had all the relevant data for running this particular regression. Flipcharts were used in three subjects--science/agriculture, math, and home science/business education. Eighteen of the 83 schools had flipcharts, 65 had none at all. And among the schools that had flipcharts, some schools had one flipchart only, and others had up to 15. Figure 2 presents the distribution of flipcharts among the 83 schools:

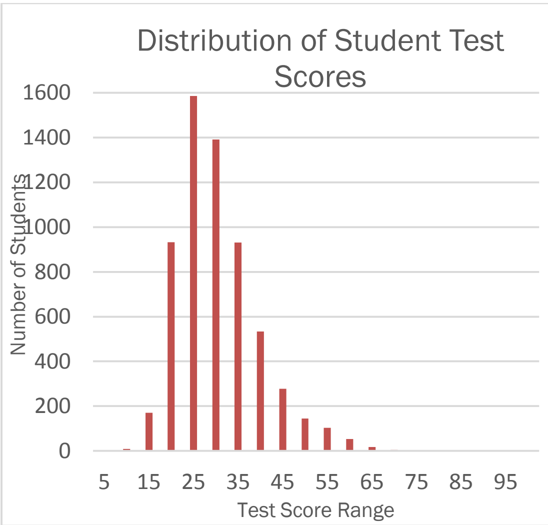
FIGURE 2



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

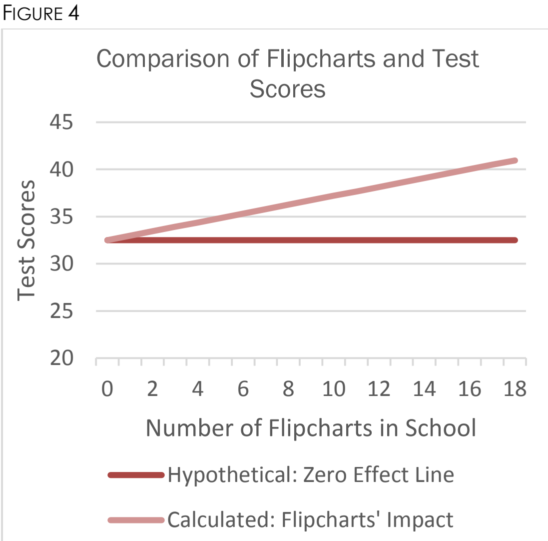
The exams covered 7 subjects: math, science/agriculture, home science/business education, English, Swahili, geography/history/civics/religion, and arts/crafts/music. Of these seven, the authors only analyzed the first 3 subjects in the initial regressions (the flipchart subjects). Figure 2 presents the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

FIGURE 3



The average test score is around 36. But with a distribution of test scores, and a distribution of flipchart availability, the question of interest was: how are these two distributions related? Do students from flipchart schools learn more over the course of the year? Figure 4 represents the relation.

FIGURE 4



The comparison line has a positive slope which illustrates a positive correlation between flipcharts and test scores. This means that students in schools with more flipcharts, in fact, tend to learn more. (The regression line suggests that students with no flipcharts seem to receive an average score in the low 30s. Students with 5 flipcharts seem to score closer to the mid-to-upper-30s range.) If there were no relationship, we would expect to see a flat, horizontal line (the dotted line).

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

This observational study found that an additional flipchart would lead to an increase in test scores of 0.47 points, suggesting that flipcharts were effective (see Table 1 or "slope" in Figure 4). But as can be seen in Table 2 below, there are other differences between schools that have flipcharts and schools that do not, apart from the availability of flipcharts. For example, schools with flipcharts were more likely to have blackboards than schools without flipcharts. So it is possible that the greater availability of blackboards was at least partly responsible for why students in schools with flipcharts did better than students in schools without flipcharts.

TABLE 1
Impact of Each Additional Flipchart on Test Scores Analyzing Flipcharts, Alone, and Controlling for Other Variables

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

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

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

TABLE 2
Distribution of Other Variables Between Schools With Flipcharts and Schools Without Flipcharts

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

Discussion Topic I
Observational Study

1. Do you believe the results of this evaluation reflect the true effect of flipcharts on student test scores for this set of schools? Why? Why not?

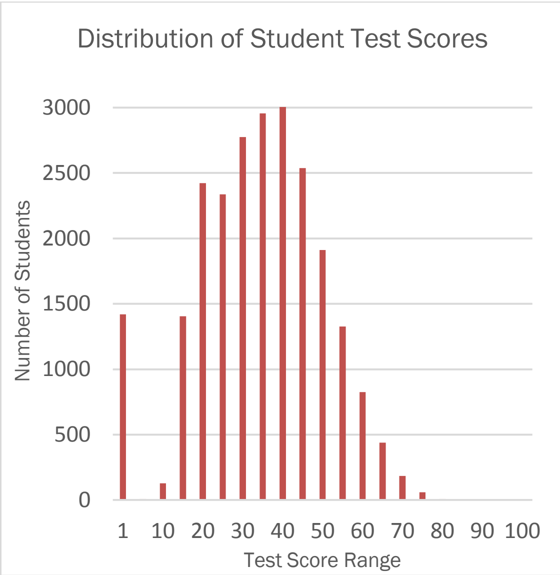
RANDOMIZED EVALUATION

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

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

Figure 5 shows the distribution of test scores after 1 year of flipcharts being in place.

FIGURE 5

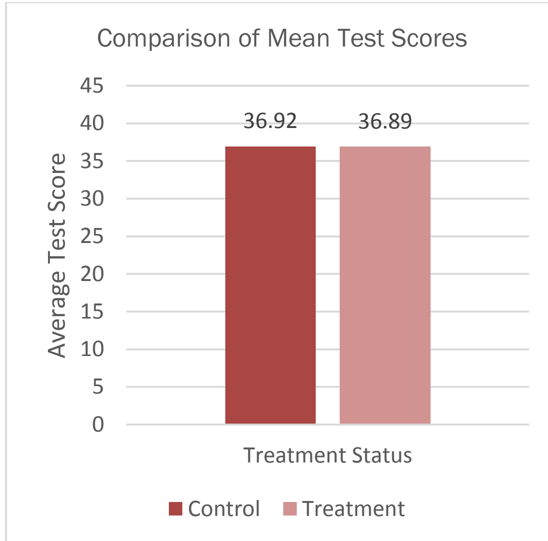


Looking at Figure 5, it is clear that, as in the retrospective evaluation, not all students were at the same level at the end of the year. The mean test score for all schools is around 36 points. But how does this distribution break up between treatment and control schools?

This question is answered in Figure 6.

Figure 6 shows the mean test scores in the treatment and control groups. Unlike the prior evaluation, in which the flipchart-to-test score graph had a line with a positive slope, therefore showing a positive relation between the two, in this figure, there doesn't seem to be much difference in test scores at all. Regardless of whether schools had flipcharts or not, their test-scores seemed to be about the same. This suggests that flipcharts had zero impact on child performance.

FIGURE 6



Discussion Topic 2

Randomized Evaluation

1. Do you believe the results of this evaluation reflect the true effect of flipcharts onstudent test scores for this set of schools? Why? Why not?

CONCLUSION

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

Discussion Topic 3

Comparison of Evaluations

2. Why do you think the results from the two evaluations are different?

	Objectives Hierarchy	Indicators	Sources of verification	Assumptions
Impact				
Outcomes				
Outputs				
Inputs				
Needs				

REFERENCES

Davis, O. L., Jr. 1968. "Effectiveness of Using Graphic Illustrations with Social Studies

Textual Materials. Final Report." Kent State University, Ohio..

Dwyer, Francis M., Jr. 1970. "Exploratory Studies in the Effectiveness of Visual

Illustrations." AV Communication Review, 235-49.

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

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

Holliday, William G. 1973. "A Study of the Effects of Verbal and Adjunct Pictorial

Information in Science Instruction." Mimeo, Ohio State University.

Samuels, S.J. 1970. "Effects of Pictures on Learning-To-Read, Comprehension, and

Attitude." Review of Educational Research, Vol. 16, 397-407.

Shepard, Roger. 1967. "Recognition Memory for Words, Sentences, and Pictures."

Journal of Verbal Learning and Verbal Behavior, Vol. 6.