Textbooks and the Paradox of Standards[1]

Paul Glewwe, Michael Kremer, and Sylvie Moulin[2]

July, 2004

Abstract

Although there is intense debate about the effect of increased expenditure on education in developed countries, there is widespread consensus that provision of textbooks can substantially increase test scores in developing countries. This paper evaluates a program through which a Dutch non-profit organization provided textbooks to randomly selected rural Kenyan primary schools. There is evidence that increased availability and use of textbooks due to the program raised test scores for those in the top two quintiles of initial academic achievement, and that it increased the probability that students would go to secondary school. However, there is no evidence that textbook provision increased test scores for students with weaker backgrounds, or reduced grade repetition or dropout rates.

[1] This project was a joint undertaking of many people. The data were collected and analyzed by Charles Asoka, Najy Benhassine, Marcos Chamon, Edward Drozd, Daniel Haar, Nauman Ilias, Dan Levy, Irene Majale, Sean May, Ted Miguel, Robert Namunyu, Caroline Nekesa, Stacy Nemeroff, Jonah Rockoff, Jaypee Sevilla, Michael Wambetsa, Polycarp Waswa, Stanley Watt, and Maureen Wechuli. Invaluable assistance was provided by the staff of Internationaal Christelijk Steunfonds: Chip Bury, Jos Huizinga, Paul Lipeyah, Japheth Mahagwa, Rebecca Mbaisi, and Susan Walji. We are grateful to Joshua Angrist, Angus Deaton, Paul Gertler, Claudia Goldin, Douglas Gollin, Eric Hanushek, Lawrence Katz, Marlaine Lockheed, Richard Murnane, Steve Pischke, and Duncan Thomas for advice and comments. Finally, we are grateful to Joash Patrice Munala, the District Education Officer of Busia, to the late George Buluma, the inspector of primary schools in Busia, and to the headmasters, teachers, and students of the participating schools. The costs of this evaluation were covered by the National Science Foundation and the World Bank research committee. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank, its Executive Directors, or the countries they represent.

[2]Glewwe, University of Minnesota and World Bank; Moulin, World Bank; Kremer, Harvard University, Brookings Institution, Center for Global Development, and NBER.

I. Introduction

Should educational standards be common for all, or should pupils be tracked into classes for students of different ability? A common system is often seen as benefiting disadvantaged groups. We argue that in some contexts this may work to the disadvantage of poor students and present evidence from Kenya.

This paper, and a companion paper, Kremer, Moulin, and Namunyu, (2003), both use randomized evaluations to shed light on the political economy of education in developing countries. Kremer, Moulin, and Namunyu (2003) argue that a combination of historical factors and political dynamics between various regions of the country and ethnic groups led Kenya to adopt an inequitable and inefficient system of education finance. This paper focuses on curriculum decisions, arguing that the decision to adopt a centralized education system made it very difficult to serve the entire population given the great heterogeneity in educational background among Kenyans. Given the historical legacy of colonial education in Kenya, and the political realities of most of Kenya's post independence history, the resulting education system was much more suitable for the most advantaged students than for the typical Kenyan student. In this setting, additional spending on curricula inputs, such as textbooks, may have limited value for the typical student.

Whereas the United States allows scope for considerable heterogeneity in the operation of schools through its system of local school boards, most countries have centralized systems. If students have very heterogeneous backgrounds, this can make it difficult to choose a curriculum that is appropriate for all students.

If, as in the context we examine, political economy factors give more power to elites and to teachers, then the curriculum is likely to be oriented towards students with a stronger

educational background than that of the typical student. In poor countries, market forces may produce even stronger skews toward weaker students in textbooks' design.

This mismatch in curriculum leads to high dropout rates as students are left behind and can no longer follow the material presented in class, and it also leads to high repetition rates, as it takes the typical student more time than is allotted to cover the existing material. Moreover, it leads some teachers to not even try to cover the assigned material in a year. If teachers are judged based on the performance of students on tests covering the material in their grade, teachers for the next grade will have incentives not to simply pick up where their predecessors left off, but instead to move on to teach the more advanced material in the official curriculum for that grade.

The model implies that curriculum materials such as textbooks will be of more use to advanced students than to typical students. We present evidence from a randomized evaluation in Kenya suggesting that this is the case.

Curricula are set very differently in the United States, in other developed countries, and in the developing world. In the United States, many decisions about curriculum and teaching are decentralized to local school districts. There is no national exam at the end of primary or secondary school designed to measure what students have learned (although some students who continue into post-secondary take tests that purport to measure their aptitude for post-secondary education). In the absence of such an exam, individual schools and school districts set their own standards regarding the level and pace of the curriculum. The twin mechanisms of parental choice of location and of local control over schools obviously have the potential to produce a match between the curricula and the needs of the population served by the school. Local school districts can add Advanced Placement courses, assign more or less challenging material, or

simply teach in a more or less sophisticated way. The education system also has considerable tracking within schools, allowing more advanced students to take more advanced material.

In many other developed countries, there is a national secondary school completion exam, and a single national education system rather than a system of local control. However, there are different types of schools. Students who cannot keep up with the pace of the academic track can switch to a vocational track, with a less demanding curriculum. For example, in Germany, at age 10, students of differing abilities and interests are placed into different forms of school: highly academic, commercial (high vocational) or general (part-time vocational and apprenticeship training). In 2001, only 43% of secondary students were enrolled in a highly academic school (Federal Statistical Office of Germany, Web Page).

Many developing countries, in contrast, have a single national system with little local control over schools. However, students who have trouble keeping up with the academic curriculum typically have little opportunity to switch into a separate track. In most African countries, for example, the percentage of secondary pupils attending a vocational school is lower than 5%. In Kenya, for example, only 1.7% attend these schools. A similar situation is found in Asia, where all but the most developed countries -China being the exception- have a very low percentage of vocational education students .

Specialization in developed countries typically occurs in secondary school, although sometimes at a fairly young age. However, students in developing countries may need differently tailored education even at the primary level. Primary-school students in developing countries are as old as many secondary pupils in developed countries. Among pupils starting grade eight in our sample, the average age is about 15, and.20% of the pupils are age 16 or older; among pupils starting grade six or seven the average age is 13.6 and 38% are age 14 or older. Second, most

students in developed countries enter some type of tracking system long before they finish their education. A typical student in the U.S. who finishes 13 years of education spends around half of his or her education in a tracked system. In contrast, a typical student in Kenya who completes eight years of education will spend none of it in a tracked system. The typical child in United States has parents with secondary education and lives in a society in which secondary education is taken for granted. The combination of parental education and general school quality is such that the typical child does not have trouble making it through the relatively standard primary curriculum and therefore there is less need for tracking at this stage. In many developing countries, this is not the case and there may be a need for different curricula to suit different students even within primary schools.

As the education system in developing countries has expanded rapidly, so has the diversity of abilities and backgrounds of students. Yet, the system has not changed accordingly. In 1960, just prior to independence, the gross primary enrollment ratio in Kenya was only 47%. By 1980, it has almost doubled to 90% (Deolalikar, 1999). This push for universal primary education enshrined in the Millennium Development Goals calls for further expanding access to education. In this paper, we argue that the failure to adjust curricula in line with this expansion has led to a situation in which the system is ill-suited for the typical student, and in which expanding the provision of such a basic input as textbooks benefits only the academically strongest pupils.

When the US education system expanded rapidly, curricula shifted as well. At the beginning of the twentieth century, the US secondary school system served only 10% percent of the population and was highly academic, preparing students for tertiary education. Even in 1910, despite the fact that high school enrollment had doubled from its level in 1900, 50% of high

school students took Latin (Ravitch, 2000). During the following decades, the United States underwent a tremendous expansion of secondary education; by mid-century, high school enrollment has increased to 65%. As part of this expansion, the curriculum was transformed (Ravitch, 1983). Under the influence of the Progressive movement, vocational and other nonacademic studies became a vital part of education, while other subjects, such as history and other humanities, were severely curtailed. Many see this as a dumbing down of the curriculum and trace the decline of US education to the educational philosophies in place at the time (See for example Finn et.al, 1984). Many countries in Africa, Kenya included, have had similar rapid expansion of education , and although there have been complaints about the decline in the quality of education in Kenya; there has been no comparable revamping of the curriculum.

Our results also have implications for the debate about the effect of education expenditure on test scores. While many have expressed skepticism about the impact of education expenditure in both developed and developing countries [Burtless, 1996; Card and Krueger, 1992; Hanushek, 1995], there is wide consensus that providing textbooks in environments where they are scarce substantially increases test scores (see literature reviews by Heyneman, Farrell, and Sepulveda-Stuardo [1978], Fuller [1986], Lockheed and Hanushek [1988], and Fuller and Clarke [1994]). Policymakers have acted in accordance with this view [Lockheed and Verspoor, 1991]. For example, when the World Bank recently scaled up lending to Kenya following the end of the Moi regime, one of the first loans was for a massive textbook supply program.

Unfortunately, most of the existing literature around which this consensus has developed is based on retrospective studies and thus may suffer from omitted variable bias. Retrospective studies will overestimate the impact of textbooks if schools with more textbooks have other, unobserved, advantages, such as parents who are more committed to education. They could also

underestimate the effect if textbooks are given to especially needy schools. For example, in 1994 the Jomo Kenyatta Foundation and the World Bank provided textbooks to Kenyan schools identified as particularly disadvantaged.

These potential biases can be avoided, at least in theory, by implementing prospective studies. The World Bank undertook one such study the early 1980's in Nicaragua [Jamison, et al. 1981]. That study found that textbooks raised test scores by about one third of the standard deviation of individual test scores.

In contrast to previous studies, this study finds that provision of textbooks to randomly selected Kenyan primary schools raised test scores only for the strongest students. Specifically, during the first two years of the program the scores of students with pre-test scores in the top 20% of the distribution increased by about 0.2 standard deviations (of individual test scores). There was no such effect for students with lower pre-test scores. Grade 8 students in the primary schools that received textbooks in the first year were more likely to enter secondary school in the following year. This is consistent with the finding of test score increases only for strong students, since only those students finish grade 8 and take the secondary school entrance examination. There is little evidence that textbooks affected grade repetition, dropout rates or attendance, which is consistent with no impact on weaker students. Textbooks also did not affect pedagogical practices, as measured through classroom observation.

The results do not appear to be statistical artifacts. The treatment and comparison schools were similar in geographic location, enrollment, and test scores prior to the program. Neither selection, nor attrition bias, nor crowding out of other sources of funding appears to drive the results, and attempts to correct for these problems do not change the findings. In addition, another program that provided grants to schools showed similar results when the grant was spent

on textbooks (Glewwe, Iyer and Kremer, 2002); scores were mostly unaffected but students with higher pre-test scores appeared to benefit.

Our results have several implications, both methodological and substantive. Methodologically, they suggest that retrospective estimates of the impact of school inputs may be misleading. Substantively, they suggest that Kenya's educational system may be so strongly skewed towards stronger students that provision of textbooks has limited value for the majority of students. The Kenyan system is currently oriented toward the top students, with curricula that are often too difficult for the average student. One solution to this is to have students complete the curriculum at different paces. This already takes place in practice, but in a manner that is highly inefficient.

This paper is organized as follows: Section II reviews the literature and provides background on primary education in Kenya. Section III describes the retrospective impact of the program on test scores. Section IV discusses the design of the program and some basic information on textbook use during the program period. In Section V we discuss the impact of the program on school attendance and grade progression, and Section VI addresses the primary question of the program effect on test scores. Section VII addresses the possibility of crowding out of the textbook program by other fundraising, and Section VIII summarizes and discusses policy implications.

II. Background

II.A. Primary Education in Kenya

Under colonial rule, Kenya's education system was designed to produce a small group of elite Africans to work in the colonial bureaucracy. It thus had a relatively demanding curriculum

without any remedial education for those who fell behind. Following independence, references to English history and geography have been replaced with references to Kenyan history and geography. Pictures in books are of Kenyan, not English, children. But the language of instruction remains English, which is most students' third language. The curriculum is arguably better suited to students with educated parents than to the typical pupil.

These historical patterns in the curricula have persisted due to the combination of great heterogeneity within Kenya, imposed uniformity of curricula, and concentration of power with those likely to prefer a system oriented towards more advanced students.

Kenya has great heterogeneity in socioeconomic status and parental attitudes toward education. Some pupils grow up in communities where parents have little education, teachers are frequently absent, and health problems, demands for child labor, and community attitudes are such that absenteeism among pupils is relatively high; whereas others grow up with educated parents, who insist on high rates of attendance among pupils and teachers.

The heterogeneity in background among Kenyan students is combined with homogeneity in curricula, as seen in Kenya's uniform curriculum, national exams, and centralized teacher training. In some educational systems, such as that of the U.S., each school can, within limits, choose textbooks and curricula that are suited for its own students. In centralized systems, such as Kenya’s, this is very difficult. The costs of this centralization may be limited in relatively egalitarian countries such as Japan, but they are likely to be high in countries with substantial ethnic, economic, and educational heterogeneity, such as Kenya.

However, even given Kenya's heterogeneity and its decision to adopt a uniform curriculum, Kenya could have adopted a curriculum and textbooks that would be targeted to the average student. However, both political economy and market factors contribute to a system in

which curricula, and textbooks in particular, are not targeted towards the typical student. The Kenyan political system has historically not distributed power very widely, and elites have had disproportionate political power. These elites are likely to prefer an educational system targeted to their children's needs. Moreover, the Ministry of Education officials in charge of designing curricula and textbooks are naturally likely to represent the interests of their children, as well as to reflect the norms of their profession, which are likely to see efforts to design a more suitable curriculum as "lowering standards." To the extent that elites in post-independence Kenya are seeing things through the prism of international comparisons, and in particular comparisons with the United Kingdom or the United States, they may resent efforts to adjust the level of the curriculum and textbooks.

Given the available exit options for the most and least advanced pupils, the median parent is likely to support designing a curriculum that is tilted towards more elite families. To the extent that elites would respond to a less demanding curriculum by leaving government schools and entering the rapidly growing private school system, the median parent may prefer a curriculum that keeps elites in the system. Conversely, to the extent that an overly demanding curriculum causes weak pupils to leave the system, typical students may benefit from better average peers, as well as from more resources per student. Finally, given that the market for textbooks consists primarily of elite students and teachers themselves, and that poorer students typically do not buy textbooks, it's of little surprise that textbooks are designed to be useful for students at the top of the distribution rather than for more typical students.

Incentives for teachers also create demand for advanced textbooks. Given that primary schools are judged by the average performance of students on the KCPE exam, they have an incentive to target their instruction to the top students in the school. This comes about because

students who drop out of school before taking the KCPE exam are not included in calculations of schools' performance, teachers have incentives to buy advanced textbooks and to use them in a way that benefits students at the top of the distribution rather than throughout the distribution.

Given that Kenya was a *de facto* or *de jure* one-party state through most of its history, but that there was a fair degree of local competition, politicians had to appeal to voters, but could not credibly argue that they could easily change national policies. They could, however, offer to bring home the bacon in terms of directing national resources to their constituencies. The political system, with its emphasis on pork-barrel politics, thus provides little incentive for politicians to devote much energy to influencing the national curriculum.

Almost all children in Kenya enroll in primary school. Students in grades one, two, and three are taught in English, Kiswahili, and the local language (Kiluhya in two thirds of our sample and Ateso in the rest). From grade four on, all instruction is in English. At the end of grade eight, Kenyan students take a national exam that determines which secondary schools they can attend. In many primary schools, only the best students are promoted to grade eight, while the rest repeat grade seven or drop out. Many students drop out before reaching higher grades; in our sample about 35% of the grade 3 students in 1996 had dropped out by 1999.

In Kenyan schools, the Ministry of Education sets the curriculum, administers national and district exams, and hires, transfers, and pays teachers. Local school committees, composed primarily of parents, must raise funds for all other costs of running the school. Fundraising for major capital expenses, such as construction takes place through *harambees* – large-scale fundraising drives. Recurrent costs, such as minor repairs, chalk, and books for teachers have historically been covered through various school fees. In practice, headmasters and parents often bargain over how much of the official fee a particular family must pay. Physical facilities at

these schools are minimal; classrooms are often dilapidated, and in some cases non-existent. Most students sit two or three to a bench, and some sit on the floor. Schools usually have textbooks for teachers, but few provide them to students. A Ministry of Education survey found a pupil-textbook ratio of 17 in primary schools in 1990.

Most schools own very few textbooks; 80% of the students in the sample are in classrooms where the school provides less than one English textbook for every 20 students, and the analogous figures for math and science are 78% and 89%, respectively. Our data show that, in grades three, four and five, only one out of six students has textbooks in English and mathematics, while in grades six and seven one out of four has these textbooks. Very few students have textbooks in other subjects. Students who reach grade eight have more textbooks; about 40 percent have mathematics and English textbooks. About 80-90% of the textbooks that students have were purchased by their parents, rather than provided by the school. Typically, the two or three students sit on a bench squeezed together, and if a book is available, they share it between them.

III. Retrospective Estimates of the Impact of Textbooks on Test Scores

The literature based on retrospective data typically suggests that, in schools with few textbooks, textbook provision can substantially increase scores. Analysis of retrospective data from Kenya yields similar results.

Lockheed and Hanushek [1988] review studies of textbooks in developing countries; the four studies examined in most detail report that textbooks raised test scores by 0.34, 0.36, 0.30 and 0.06 standard deviations of individual test scores. Other literature reviews cast a wider net, perhaps including studies of lower quality. Heyneman, Farrell, and Sepulveda-Stuardo [1978]

find positive effects of textbooks on test scores in 15 of 18 studies, and note that studies typically find greater effects of textbooks among disadvantaged students. Fuller [1986] reports significant effects of textbooks in 14 of 22 studies, and Fuller and Clarke [1994] found significant effects in 19 of 26 studies. The only randomized evaluation we know of that examines the impact of textbooks on student performance is that of Jamison, et al. [1981], which also examined the impact of radio education.[3] That study compared 48 first-grade classrooms in Nicaragua that were randomly assigned to receive radio mathematics education, with 20 that received mathematics workbooks and 20 that served as controls. After one year, pupils who received workbooks scored one-third of a standard deviation higher than the control group pupils, a difference that was significant at the 1% level. Pupils in radio mathematics schools scored more than a full standard deviation higher than control group pupils. No significant interaction was found between pre-test scores and the textbook intervention, although textbooks narrowed gaps between rural and urban students.

Cross-sectional variation in textbook availability in the 75 schools in our data from Kenya that did not receive textbooks in the first year of the program examined in this paper were used to estimate the relationship between textbooks and test scores using non-experimental data. The tests used, which were administered to students in grades 3-8 in October and November of 1996, covered English, math and science. (These and other tests used in this paper are explained in detail in Section IV.) Because test scores are measured in arbitrary units, for each grade and subject combination we normalize test scores by subtracting the mean test score in these 75 schools and then dividing by the standard deviation. Thus a student who scored 0.1 was 0.1

[3] Heyneman, Jamison and Montenegro [1984] compare Philippine schools randomly selected to receive either one or two textbooks per student, and they find little difference in test scores. They also present results comparing these two scenarios with no textbooks, which suggest a substantial positive impact of textbooks on test scores. Yet these estimates compare the same schools before and after receiving textbooks and so are not based on randomized trials.

standard deviations above the mean score. For reference, an increase from 0.0 to 0.1 standard deviations in a normal distribution moves a student from the 50th percentile to the 54th.

For these 75 schools we used least squares (with school-level random effects) to regress the (normalized) test scores of students in grades 3 though 8 on (i) a dummy variable indicating that a student has a privately owned textbook; (ii) the textbook to student ratio for school-owned textbooks (calculated separately for each grade/subject combination in each school); and (iii) other school and family characteristics.

Students with privately owned English textbooks scored 0.18 standard deviations higher on English exams, controlling for parental education and land ownership. Those with math and science books scored 0.09 and 0.05 standard deviations higher on the exams in those subjects (Table 1). The English and math impacts are highly significant, while the smaller science impact is statistically insignificant. Aggregating over all three subjects yields an impact of 0.12, with a tight standard error of 0.016.

In contrast, school-owned textbooks are negatively correlated with test scores in the OLS specification. Students in schools that own more English textbooks have slightly lower English test scores, but the relationship is far from significant. School-owned math textbooks have a large and *significantly* negative coefficient (-0.35). Finally, students score 0.55 standard deviations lower in science tests in schools that own an additional textbook per pupil. Aggregating across all subjects, test scores are -0.19 standard deviations lower in schools that own an additional book per pupil. Most of the other variables are not statistically significant (results not shown) but those that are yield sensible results; for example both father's post-secondary education and household wealth are associated with higher test scores.

One potential reason for the negative correlation between school-owned textbooks and test scores is that the government or donors may provide more textbooks to the neediest schools. One method to address the resulting omitted variable bias is difference-in-differences estimation. A World Bank textbook distribution project carried out by the Jomo Kenyatta Foundation provided textbooks to 95 of Busia's and Teso's 334 primary schools in 1994, at ratios of about one textbook for every two pupils in English and math and one for every four in Swahili and science.

Retrospective estimates based on examining changes in test scores from 1993 to 1995 suggest that the program increased test scores substantially. School-level data on average scores on government exams in grades 6, 7, and 8 are available for about 80% of the 334 schools. In each year the test scores are normalized to have a mean of zero and a standard deviation of one (based on the distribution of student test scores, not the distribution of school mean test scores).[4] A simple difference-in-differences analysis of the data suggests that textbooks raised grade 7 test scores by 0.50 standard deviations in 1994 and raised both grade 7 and grade 8 scores by about 0.65 standard deviations in 1995 (Table 2). No significant impact is seen on the grade 6 scores, nor on the 1994 grade 8 scores.

In summary, retrospective estimates of the impact of textbooks on test scores based on cross-sectional data from Kenya indicate that privately-owned textbooks raise scores, while school-owned books reduce scores. A more sophisticated longitudinal analysis finds positive effects from the distribution of school-owned textbooks, which suggests that cross-sectional

[4] We convert standard deviations of school mean test scores into estimated standard deviations of individual test scores using a small sample of schools for which we have student level data.

estimates may be biased, perhaps because books were systematically provided to weaker schools (which indeed was the stated intention of the Jomo Kenyatta Foundation).[5]

Of course, the positive retrospective estimates of the effect of privately-owned textbooks, controlling for family background, and the positive difference-in-difference estimates of the effect of school-owned textbooks could both be biased. Parents who are more interested in their children's education, a characteristic that is difficult to observe, may be more likely to purchase textbooks, or parents may consider textbooks to be complements to their children's ability, and thus provide textbooks to more able children. Schools that received textbooks from the Jomo Kenyatta Foundation may also be better placed politically to receive other inputs or may have recently been assigned more able headmasters. These possibilities suggest that we should examine results from a randomized evaluation.

IV. The School Assistance Program

In late 1995, the Ministry of Education district office selected 100 of the 333 primary schools in Kenya's Busia and Teso districts to participate in the School Assistance Program (SAP) funded by International Christelijk Steunfonds (ICS), a Dutch non-profit organization. The 100 SAP schools were chosen because they were thought to be particularly in need of assistance and (with one exception) had not participated in the Jomo Kenyatta program discussed in Section III. The median school average test score among SAP schools on the grade 6 and 7 district exams in 1995 was at the 40th percentile of the distribution of school average test scores of all Busia's schools. On the grade 8 exam, the median SAP school was at the 33rd percentile.

[5] Test score data, however, show little evidence that weaker schools were chosen among schools receiving textbooks from this project, the median test scores of the 6th, 7th, and 8th grades on the 1993 district exam were at the 54th, 48th and 48th percentiles, respectively, of all primary schools in the district.

The 100 SAP schools were randomly divided into four groups. Schools were listed according to alphabetical order within geographic divisions. From this list, every fourth school, starting with the first, was assigned to group 1. Similarly, every fourth school starting with the second, third, and fourth was assigned to groups 2, 3, and 4, respectively. Group 1 schools received textbooks in early 1996. In early 1997, group 2 schools received a grant that could be used for a variety of educational material (including textbooks). As discussed below, group 2 schools spent a large proportion of the grants on textbooks, while group 3 schools spent more heavily on construction. Schools in groups 3 and 4 received similar grants in early 1998 and early 2000, respectively. In the rest of this paper, 1996, 1997, 1998 and 1999 are referred to as years 1, 2, 3 and 4, respectively, of the program.

Appendix Table A.1 summarizes the test score data used in this paper. The Ministry of Education administers to upper-grade students district-wide exams on all subjects in the Kenyan curriculum. These achievement tests are designed to measure student understanding of the official government curriculum. Grade 8 students take the Kenya Certificate of Primary Education (KCPE) exam, which determines continuation into secondary school.[6] ICS administered additional tests modeled closely on the Kenyan government tests and developed by Ministry of Education officials over four years in the 100 schools it assisted. First, in January 1996 (the Kenyan school year runs from January through November) ICS administered baseline tests in English, math, and science for grades 3 through 8 in the 25 schools that received textbooks and the 25 schools that were not assisted until 2000 (the other 50 schools were assisted

[6] In 1996, district exams were given in grades 5 through 8. In 1997 they were given only in grade 8, due to a national decree unrelated to the textbook program. In 1998, they were given in grades 4 through 8 in Busia district, but only in grade 8 in Teso district. In 1999 these exams were given in grades 4 through 8 in both districts. District exams are given in October for grades 4 through 7 and in July for grade 8. The KCPE is given in November. Unless otherwise stated, district exam results use the October test for grades 4 through 7 and the KCPE for grade 8.

in 1997 and 1998, as explained in Section V). Then, for all 100 schools, ICS supplemented the October 1996 district exams in grades 5 through 8 with tests in grades 3 and 4. In October 1997, ICS supplemented the district grade 8 exam with tests in grades 3 through 7. In October 1998, ICS conducted exams in grades 3 through 8 even though there were district exams in grades 4 through 8 in Busia district and in grade 8 in Teso district. Finally, in October 1999 ICS again administered exams for grades 3 though 8.

The 25 schools that received textbooks in year 1 can be compared to several sets of comparison schools. At the end of year 1, they can be compared to all schools that had not yet received assistance: groups 2, 3, and 4. We call these schools the 75-school comparison group. At the end of year 2, they can be compared to the 50 schools that had not yet received assistance: groups 3 and 4. This is the 50-school comparison group. In years 3 and 4, they can be compared to the 25 group 4 schools that were not assisted until early 2000; this is the 25-school comparison group. Any results that use the January 1996 pre-test scores must compare the textbook schools with the 25-school comparison group, since only those comparison schools have pretest scores.

IV.A. Initial conditions in textbook and comparison schools

The 25 textbook schools and the 25-school comparison group had very similar pre-program test scores (Table 3). As in the retrospective analysis, for each grade and subject combination we normalize all test scores by subtracting the mean test score in the comparison schools and then dividing by the comparison school standard deviation. All regression estimates use school random effects, since students within a school may be subject to common shocks, for example, due to headmaster quality.

Averaging across grades, pre-program differences in English and math test scores are never more than 0.05 standard deviations and never statistically significant. The same is true for science when averaging over all grades, but the difference for grade 8 in science (the only grade that received science textbooks) is large, 0.17, although statistically insignificant. Results at the bottom of Table 3 from regressions that stack over all grade/subject combinations show small and statistically insignificant differences: 0.06 standard deviations for grades later provided with textbooks and 0.02 for all grades.

IV.B Textbook Provision

ICS provided the 25 Group 1 schools with official government textbooks, published by the Kenya Institute of Education, in February 1996 (the second month of the Kenyan school year). Grades 3 through 7 received English textbooks, and grades 3, 5 and 7 received math textbooks. Grade 8 received science textbooks since almost half of those students already had math and English textbooks. In January 1997, math textbooks were given to grades 4 and 6, and agriculture textbooks to grade 8. For each grade and subject that received textbooks, ICS also provided one copy per grade of the associated teacher's guide.

Textbooks were given at less than a one-to-one ratio, based on Heyneman, Jamison, and Montenegro's [1984] finding of little difference in test scores between Philippine schools randomly allocated one textbook for every two pupils and those allocated one textbook for each pupil. Sharing textbooks is common in rural Kenyan primary schools, where two or three students typically share a bench and desk. A 60 percent textbook per pupil ratio was provided in English and science, and a 50 percent ratio in math.

In year 1, the program greatly increased the supply of textbooks in the textbook schools relative to the supply in comparison schools. However, the impact declined over time as books depreciated, comparison schools obtained books from other sources, and the program crowded out purchases of new private books in the textbook schools. In the first year of the program, the ratio of school-owned books per pupil for grade-subject combinations that received textbooks is 0.65 in textbook schools but only 0.04 in the comparison schools (Table 4). In contrast, in the grade-subject combinations that did not receive textbooks the ratios are virtually identical for both groups of schools, 0.02 and 0.03. In grade-subject combinations that received textbooks, textbook schools have fewer privately-owned textbooks than comparison schools, 0.09 vs. 0.15, suggesting that the program crowded out 0.06 private books per student in year 1. Yet such crowding out is small compared to the difference in school-owned books; combining school and private books (columns 5 and 6), the ratio is 0.74 for the textbook schools and 0.19 for the comparison schools, a gap of 0.55. In contrast, in the grade-subject combinations that did not receive textbooks the ratios were almost identical, 0.08 and 0.09.

Similar results hold in later years, but the gap between the textbook and comparison schools narrows over time in grade-subject combinations that received textbooks. In year two, the average book per pupil ratio for school-owned books falls to 0.55, perhaps due to loss of (or damage to) school-owned textbooks. Crowding out appears to increase slightly; privately-owned books per pupil were 0.09 and 0.17 in the textbook and comparison schools, respectively.

At the end of year 2, under a new program financed by the Dutch government, the district education authorities distributed textbooks to 21 of the 100 schools. (These 21 schools were spread evenly over the four groups of 25 schools.) Thus, in year three the stock of school-owned books held steady in textbook schools while the comparison schools increased their stock. This

further narrowed the gap in textbook stocks between the groups, and also meant that some of the textbooks in textbook schools may have been duplicates, further narrowing the gap in the effective textbook ratio. Moreover, there were some curriculum changes, and new editions of some textbooks appeared at the beginning of year 2. While the changes were modest, education in Kenya is structured around the official curriculum, and some teachers may have obtained new textbooks and been reluctant to have their students use an older version of the text than the one they were teaching from. Finally, by year 4 the gap in textbooks had declined even further for grade-subject combinations that received textbooks.[7] For school-owned books the ratio was 0.43 in textbooks schools and 0.10 in comparison schools, a difference of 0.33. When private books are included, the ratios are 0.48 and 0.21. The provision of additional textbooks to some schools at the end of year 2, and to a lesser extent the change of some textbooks and curriculum in year 2, lead us to focus most of our analysis on the results from years 1 and 2.

IV.C. Textbook Use and Pedagogy

Data on actual textbook use and pedagogy come from two sources. The first is a student questionnaire administered in years 2 and 3 to students in grades 6-8 (students in grades 3-5 were considered too young to fill out this questionnaire). The second is classroom activity data collected by trained observers who recorded minute by minute information on teacher and student activities. Six conclusions can be drawn from these data.

First, students in textbook schools were more likely to have opportunities to take school-owned textbooks home to study. In year 2, 53% of textbook school students report being able

[7] The year 4 figures in Table 4 are for English and math books only, and include only students in grades 6-8. Data are not available for science books in year 4, and no data were collected for pupils in lower grades in year 4.

to take home school-owned textbooks in the subject-grade combinations that received textbooks. Only 3% did in comparison schools (Table 5).

Second, the lower pupil-textbook ratio in textbook schools led to more access to textbooks in class. Sixty-two percent of textbook school students in year 2 report having access to a school-owned book in class for the grade-subject combinations that received textbooks, compared to only 8% of students in the 50-school comparison group (Table 5). In grade-subject combinations that did not receive textbooks only 9% of students had textbooks in textbook schools, almost the same as in comparison schools (7%).

Third, it appears that comparison schools make use of the few textbooks they have by having students share them in relatively large groups (Table 6); as explained above, in Kenyan classrooms the seating of students on benches easily allows two or three students to share a textbook. In year 2, about 44% of textbook school students either had their own copy or shared a copy with another student. In contrast, only about 20% of comparison school students had their own textbook or shared with one other person. 26% of students in textbook schools had no textbook at all or shared a textbook in a group of 4 or more students. In the comparison schools, 51% had no textbook or shared in a group of 4 or more. A similar pattern is observed in year three.

Fourth, the difference in textbook use between textbook schools and comparison schools narrowed between years 2 and 3. For example, whereas only 8% of students in comparison schools had been issued a textbook for use in school in year 2, 28% had by year 3 (Table 5). Interestingly, by year 3 textbook schools have more school-owned textbooks than the comparison group in grade-subject combinations where ICS did not provide textbooks,

suggesting that ICS textbook provision allowed schools to focus new purchases on other grade-subject combinations.

Fifth, pupils used textbooks more in class in textbook schools, although differences were modest and had largely dissipated by year 3. It may be that teachers were more comfortable using textbooks when all students had access to them. This increase in usage occurred on two margins. First, teachers in textbook schools were more likely to be present in school, at least in year 2. Absence was 38% in the comparison group, but only 24% in the textbook schools (Table 6).[8] This did not persist to year 3. Second, teachers were more likely to use the textbooks while in class. In year two textbook school students used textbooks in 62% of the classroom sessions observed, compared to 46% in the 50-school comparison group.[9] This difference, which includes sessions where teachers were absent (in which case students are unlikely to use textbooks), is significant at the 5% level (Table 6). Conditional on the teacher being present, in year 2 textbooks were used in 79% of the observed sessions in the textbook schools, compared to 68% in the comparison schools, a difference significant at the 10% level. In year 3, the unconditional difference is smaller, 46% vs. 37%, mostly because of little difference in teacher absences. The difference conditional on the teacher being present is similar, 64% vs. 52%, and statistically significant at the 5% level. An interesting difference between the two years is a decline in the use of textbooks in both types of schools; this may reflect the fact that new editions

[8] Classroom observation data were not collected in year 1. There are classroom observation data for year 4, but because those data did not focus on the use of English and mathematics textbooks the sample sizes for those two subjects are too small (only 23 observations, compared to 222 observations in year 2 and 165 in year 3).

[9] The numbers on use of textbooks in class for comparison schools in Table 6 at first appear quite high compared to the numbers on school owned textbooks in Table 5. Yet there are two mitigating factors. First, some students have privately purchased textbooks, as seen in Table 4. Second, the figures in Table 6 are in response to a general question: "Did the pupils use the textbooks during this lesson?' The response to this question is "yes" if this occurred for students with textbooks, even though students without them could at best borrow other students' books.

came out for some textbooks, but it certainly is not due to general textbook availability, which hardly changed from year 2 to year 3 (Table 4).

Sixth, students in textbook schools report doing more days of homework per week (Table 6), which is plausible because the availability of textbooks for students to take home provides a source of homework assignments for teachers. Even so, the increase in the number of days is not particularly large – in year 2 students in textbook schools report doing homework 4.7 out of the last seven days while students in the comparison schools report 4.1 days. The analogous numbers for year 3 are 4.2 and 3.8, respectively.

IV.D. Crowding Out

There is some evidence that the program crowded out other fundraising by schools, but these results are statistically insignificant and are limited to smaller schools. These findings are based on data from a school questionnaire and a school committee questionnaire. The 1997 school questionnaire asked about funds collected from all local fundraising events in 1996, and about assistance from NGO's and other donors. The 1997 school committee questionnaire asks only about funds raised from harambees, the most common local fundraising event, which were discussed in Section II.[10] These events provided one third of the funds raised in the 75 comparison schools in 1996, but are relatively infrequent; of those 75 schools, only 15 had Harambees in 1996. In those 15 schools, Harambees provided 72 percent of all funds raised, almost all of which were used for construction.

[10] School finance data are primarily from the school questionnaire. However, if funds from harambees reported in the school committee questionnaire exceed funds from harambees reported in the school questionnaire, the amount reported the former questionnaire is used since that questionnaire collected more detailed information on harambees.

The assistance (in U.S. dollars) received from ICS and from other sources in 1996 differed for the textbook and comparison schools (Table 16). The average value of textbooks from ICS was $485. The average amount of non-ICS aid was $456 for the comparison schools and $267 for the textbook schools. The difference of $189 suggests that receipt of ICS textbooks crowded out other assistance by an amount equal to about 39% of the assistance from ICS.

These differences in non-ICS assistance, while large, are statistically insignificant. The high standard errors reflect the large variation in non-ICS assistance; most schools received nothing, while a few received large amounts. Crowding out of other fund-raising is unlikely to explain the lack of a program effect on average test scores. First, even if ICS assistance reduced harambee fundraising, there would be little short-run impact on educational outcomes because constructing new facilities takes time and the flow of services from new facilities extends over many years. For example, if providing a dollar of textbooks reduced funds raised for new classrooms by forty cents, and half of those classrooms were built in the year the textbooks were given, and the flow of services from classroom construction lasts five times longer than services from textbooks, that provision of one dollar in textbook services would lead to a four cent drop in classroom services in the first year. Thus, a 39% drop in construction funds would have little impact in the first year of the program. Second, the difference in non-ICS assistance between the textbook schools and the 75 comparison schools reflects differences among smaller schools (those with less than the median enrollment). This is plausible because smaller schools are most likely to need additional classrooms. In larger schools, where there is no evidence of crowding out, the estimated average effect of the program across all grades and subjects is 0.02, with a standard error of 0.12, which is identical to the estimate of 0.02 for 1996 in Table 9. Thus we

conclude that crowding out cannot explain the low impact of textbooks on average scores found in Section VI.

It should be noted that it is possible that by year 3 some construction could have taken effect in the comparison schools after harambees in the first year. This type of crowding out could lead to smaller program effects in years 3 and 4.

V. Program Effects on Attendance, Grade Progression, and Dropout Rates

This section argues that the program had little effect on student attendance, dropout rates, or grade repetition. However, it did increase the likelihood that eighth graders in year 1 were enrolled in secondary school in year 2 (although this effect is not seen in later years). These results suggest that the program had little effect on average and low-ability students (who are more likely to drop out or repeat a grade), but benefited high-ability students (since only high-ability students have the opportunity to enroll in secondary school).

Students enrolled in lower grades (3-7) at the start of the program could experience four possible outcomes by the second year: stay in school and be promoted, stay in school and repeat, drop out, or transfer out. The same is true for students in grade 8, except that promotion takes two distinct forms: finish primary and leave school, or finish primary and enter secondary school. After one year, children in grades 3-7 show no significant differences in dropout, transfer or repetition rates across the textbook and comparison schools (Table 8, columns 1 and 2). (Statistical significance is measured using estimates from a probit regression with school random effects). However, grade 8 students in the textbook schools in year 1 were more likely to enter secondary school in year 2 than students in comparison schools (43% vs. 38%). This difference is statistically significant at the 5 percent level and is consistent with the hypothesis

that textbooks are most helpful to the best students, since only those students reach grade 8. There is no evidence that the program affected promotion to secondary school after year 1.

The program effect on attendance is also insignificant. Since official attendance data are not very reliable in Kenya's rural primary schools, special visits were made to the 100 SAP schools to collect attendance data. The schools were visited once in year 2, four times in year 3 and three times in year 4. Averaging over all grades, attendance rates were slightly higher in textbook schools in two of the three years, but all these differences are statistically insignificant (last row of Table 8).

VI. Program Effect on Test Scores

The primary outcome considered in this paper is student test scores. This section presents a variety of estimates indicating that textbook provision raised test scores only for the strongest students. Sub-section VI.A presents three different estimates of the effect of textbooks on average test scores: a levels estimator that compares post-test scores across textbook and comparison schools, a difference-in-differences estimator comparing changes in test scores across textbook and comparison schools, and a differences-in-differences subject-based estimator that compares the difference between test scores in textbook schools in subject-grade combinations in which textbooks were and were not given to the same difference in comparison schools. All three estimation methods show little effect of textbooks on average test scores. This contrasts sharply with retrospective studies; for example, those reviewed by Lockheed and Hanushek [1988] typically found that textbooks raised test scores by about one third of a standard deviation. Subsection VI.B examines whether the impact of textbooks is related to

initial student performance. It finds that textbooks initially raised text scores by about 0.2 standard deviations for the top 20% of students, but this impact disappears in later years.

VI.A. Estimates of the Program Effect on Average Test Scores

Test scores are likely to be correlated among students in the same class and school due to unobservable teacher and headmaster characteristics, so we use an error components econometric model with school, grade, and subject random effects. Failure to account for these correlations would yield incorrect the standard errors of the program's estimated impacts. More specifically, let the test score of student i in grade j, subject k, and school s, t_{ijks}, be given by:

$$t_{ijks} = \alpha_{jk} + \beta_{jk} p_s + u_{jks} + e_{ijks} \quad j = 3,4,...8 \quad k = \text{English, Math, Science.} \qquad (1)$$

The dummy variable p_s equals one if school s received textbooks and zero if it did not. Random assignment of textbooks to schools ensures that $E[p_s u_{jks}]$ and $E[p_s e_{ijks}]$ both equal zero. All estimates of (1) use Generalized Least Squares (GLS) to account for within-school correlation of students' test scores for a given grade and subject without imposing a specific distribution (e.g. normality) on the residuals.

We also combine several grades in a given subject to estimate the (weighted) average impact of textbooks on test scores in that subject. Conceptually, there is little difference between aggregating potentially disparate effects across grades and aggregating potentially disparate effects across students, as is routinely done. This specification decomposes u_{jks} in (1) into a school specific term u_{ks} and a grade specific term conditional on the school term, v_{jks}:

$$t_{ijks} = \alpha_{3k}D_{3i} + \alpha_{4k}D_{4i} + \ldots + \alpha_{8k}D_{8i} + \beta_k p_s + u_{ks} + v_{jks} + e_{ijks} \quad k = \text{English, Math, Science.} \qquad (2)$$

Equation (2) includes grade specific intercept terms, with corresponding dummy variables.

Similarly, we present estimates that combine more than one subject in the same grade, decomposing u_{jks} in (1) into a school specific term u_{js} and a grade specific term conditional on the school, v_{jks}, to measure the average impact of textbooks in a given grade j:

$$t_{ijks} = \alpha_{jE}D_{Ei} + \alpha_{jM}D_{Mi} + \alpha_{jS}D_{Si} + \beta_j p_s + u_{js} + v_{jks} + e_{ijks} \quad j = 3, 4,\ldots 8. \qquad (3)$$

Finally, we estimate regressions that combine all grades and subjects:[11]

$$t_{ijks} = \alpha_{3E}D_{3Ei} + \alpha_{3M}D_{3Mi} + \alpha_{3S}D_{3Si} + \ldots + \alpha_{8E}D_{8Ei} + \alpha_{8M}D_{8Mi} + \alpha_{8S}D_{8Si} + \beta p_s + u_s + w_{js} + v_{jks} + e_{ijks}. \quad (4)$$

Equations (2) – (4) are also estimated using GLS to account for their error components structure.

For each year, the sample comprises all students tested in October of that year who were enrolled in January 1996 in either the 25 textbook schools or the relevant comparison group. Children who changed schools after January 1996 are always classified according to the school that they were in initially, so the estimated program effect is the impact of being offered the treatment (intention to treat), as opposed to the impact of treatment itself (issues of selection and attrition bias are discussed in detail later in this section). The comparison schools for the first year are those in the 75-school comparison group; for year two we use the 50-school comparison group, and for years 3 and 4 we use the 25-school comparison group. The smaller sample sizes for years 3 and 4, and some changes in curriculum and textbook availability that began in 1997 (explained further below), lead us to focus the discussion on the first two years of results.

The estimated program effects after one year range from zero for English to 0.06 for math (Table 9) [12]. Aggregating over all subjects gives a statistically insignificant impact of 0.02

[11] Equation (4) is appropriate if students have the same teacher for all subjects, so that w_{js} is a teacher-specific effect and v_{jks} is a subject-specific effect conditional on having that teacher. In upper grades, teachers specialize by subject, so the error term should be $u_s + w_{ks} + v_{jks} + e_{ijks}$, where w_{ks} is a teacher specific effect for the teacher teaching subject k in all grades and v_{jks} is the grade specific impact of that teacher. In practice, these two different error structures for equation (4) yield similar results. Also, adding an individual level random effect when stacking across subjects for equations (2) and (4) had almost no effect on the estimates and only slightly reduced estimated standard errors.

[12] More disaggregated results are given in Tables A4-A7 in the appendix All regressions include controls for sex and for Busia's seven geographic divisions. Regressions without such controls are also consistent and yield similar

standard deviations. The standard error of 0.086 allows one to reject (at the 5% significance level) the hypothesis that the true (average) effect was 0.20 standard deviations or higher.

There is little evidence that the impact of textbooks increases over time.[13] At the end of the second year, all textbook school students in grades 4-7 had had English textbooks for two years, and those in grades 4 and 6 had had mathematics textbooks for two years. There is little discernable impact for either subject; the point estimates of 0.06 for English and –0.06 for math are both insignificant (Table 9, panel 2). Aggregating over both subjects yields a statistically insignificant impact of 0.02, and aggregating across all subject-grade combinations with textbooks for either 1 or 2 years to maximize the sample size (24,156) yields an impact of 0.03, with a standard error of 0.09. This rules out an average impact of 0.21 or higher. All point estimates of program impact for years 3 and 4 are slightly negative, ruling out impacts of 0.20 or higher at a 5% significance level.

Comparing differences in pre-test and post-test scores across textbook and comparison schools yields point estimates of the program effect that are close to zero and more precisely estimated (Table 10). Difference in difference estimates using the schools that participated in the January 1996 pre-test, aggregating over subjects, yield point estimates for the treatment effect after one and two years of 0.02 and 0.01, respectively, and the associated standard errors rule out impacts of 0.13 and 0.11 or higher. Point estimates after three and four years of the program are slightly negative, and rule out impacts of 0.10 and 0.08, respectively.

---

results, but adding these controls often improves the precision of the estimates.

[13] Another reason to examine the results in later years is that the district tests in year 1 were quite difficult for many students; in a few cases the average scores were not much higher than could be obtained by random guessing, which implies very little information content in those tests. The ICS tests used in year 2 and later years were intentionally designed to be less difficult; the mean scores were much higher and thus they conveyed more information.

Two types of selection could lead to bias in the estimates presented above. First, the program was announced in December 1995 and early January 1996, raising the possibility that better than average, or perhaps more marginal, students enrolled in program schools at that time. Yet differences in dropout rates and transfer rates from 1995 to 1996 across textbook and comparison schools show small, statistically insignificant differences (Table 11, first two rows).

Second, textbook schools may have promoted more students in the hope that they would receive more textbooks. Twenty-two percent of textbook school students repeated a grade from 1995 to 1996 while 26% did so in the comparison schools. For a given grade in a textbook school, promoting marginal students has two effects: it adds lower performing students from the grade immediately below (who otherwise would not be promoted) to that grade and it removes students who would have been in that grade had they repeated. The first effect clearly biases downward the estimated impact of textbooks. The second effect would also lead to downward bias; data from the 75 comparison schools show that, conditional on current grade, repeaters usually perform slightly better than non-repeaters and this is particularly likely to be the case for repeaters on the margin of promotion.

However, any bias due to differential repetition is probably very small. To see why, we impose the extreme assumption that the "extra" 4% of students promoted in the textbook schools are the most marginal students: those who scored the lowest on the January 1996 tests. We then "demote" these marginal students down one grade and re-estimate the level regressions in Table 9 for year 2.[14] This yields an estimated upper bound on the aggregate textbook effect of only

[14] In 1997 each grade took a different test, so the test scores of these demoted students must be converted to an equivalent score for the lower grade to which they have been assigned. This was done using a special data set in which a sub-sample of students took the test for their grade and for one grade lower. This special data was not collected in 1996, which is why these results are only for 1997.

0.044 standard deviations, with a standard error of 0.098. This estimate is only slightly larger than the year 2 estimate in Table 9, despite using an extreme assumption.

It is also possible that the results are biased downward by attrition bias. Students who attrit are not tested,[15] so the issue is whether there are differences in the proportion of students tested (which also reflects differences in taking the tests conditional on being in school) and, if such differences exist, whether not being tested is correlated with initial academic performance. In year 1, there is almost no difference in the percentage of students not tested: 26.0% in the textbook schools and 26.3% in the 25-school comparison group (Table 11).[16] In year 2, 33.3% of students in comparison schools were not tested while 31.0% were not tested in the textbooks schools. This small difference is statistically significant, and it could lead to bias in the year two test score results if weaker students were more likely to be tested in textbook schools. However, differences in pre-test scores of students who did not attrit in textbook and comparison schools are not significant, and have different signs for different subjects. Given this, the slightly higher propensity of children in textbook schools to be tested in year 2 is unlikely to have substantially biased estimates of the impact of textbooks on test scores. In year 3, students in the comparison schools were slightly more likely not to be tested than were textbook school students (39.9% vs. 38.7%), yet this difference is not statistically significant. This is also seen in year 4; while comparison schools students are somewhat more likely not to be tested (42.9% vs. 45.2% in textbook schools) the gap is statistically insignificant. On the whole, attrition bias seems unlikely to have affected our results.

[15] We attempted to test a random subsample of dropouts, but it was very difficult to find them. For example, in 1997 only 28% of the targeted subsample were tested, and they are unlikely to be a random sample of dropouts.
[16] The other 50 schools were first visited in October 1996. At that time data were collected only for children being tested, not for children who may have dropped out or transferred out between January and October of 1996.

Despite the fact that the expected bias seems to be small we present results from a third estimator that should not be susceptible to either type of bias discussed above. This estimator is a difference-in-difference subject-based estimator that regresses test scores on dummy variables for whether students were in textbook schools *and* whether they were in subject-grade combinations that received textbooks. To understand the intuition, first think of calculating the difference in normalized test scores between treatment and comparison schools in subject-grade combinations in which textbooks were provided. Then calculate the difference in test scores between treatment and comparison schools in subject-grade combinations in which textbooks were not provided. Now take the difference between these differences; this is the subject-based estimator. Comparing subjects within a single grade that vary in whether a textbook was received effectively compares test scores in different subjects for the same student, so differences in student attrition across textbook and comparison schools will not affect the estimator (although bias could occur if students who are particularly good in a given subject are less likely to drop out if given a textbook in that subject.) In addition, selection bias as discussed above is also unlikely to affect these estimates.

In addition to avoiding bias, this approach can distinguish the effect of textbooks from the effect of other factors that affected school performance disproportionately in the treatment schools. For example, some treatment schools might have coincidentally been assigned good new headmasters. Of course, in expectation headmaster quality should be similar across both sets of schools, but this approach controls for random differences in headmaster skill. If provision of textbooks in one subject affects scores in another through spillovers we could see a positive coefficient on the textbook-school dummy. If textbooks are truly having an effect, however, the interaction between textbook subject-grade and textbook school should be positive

(except in the implausible case that textbooks systematically have larger effects on test scores in subjects other than those for which they were written).

This technique works best for the first year because that year has the most variation across grade-subject combinations in receipt of textbooks (since grades 4 and 6 did not receive math textbooks until year 2). Combining all grades and subjects into one regression (Table 12, last row) yields a slightly negative (–0.01 standard deviations) estimate of the direct effect of receiving a textbook. The small standard error of this estimate allows us to reject (at the 5% level) the hypothesis that textbooks raise average test score by 0.07 or more standard deviations, providing further evidence that textbooks did not raise average test scores.

Thus, all the estimates we examine yield point estimates of the textbook impact close to zero; depending on the strength of assumptions one is willing to impose, it is possible to reject effects as large as 0.2, 0.1, or 0.07 standard deviations. This contrasts sharply with non-experimental results, either in the literature or in our estimates for Kenya.

VI.B. Program Effects Disaggregated by Initial Test Score

Further analysis suggests that the provision of textbooks raised the test scores of students who started out with high levels of academic achievement. In the first year, an interaction term between the program variable and the pre-test score is highly significant when aggregating across all subjects (Table 13). The same is true after two years, although the interaction term is slightly lower. After three and four years the interaction effect is insignificant, with point estimates very close to zero.

Of course, pre-test scores may measure initial learning with substantial noise. The ensuing attenuation bias implies that the coefficients on the pre-test and on the interaction term

between the pre-test and the program seriously underestimate the true pre-test and interaction effects. Suppose that the true coefficient on initial academic achievement is one, as in standard difference-in-difference specifications. If so, attenuation bias is substantial, since the coefficient on the pre-test score is only 0.43 in the first year, and even lower in later years. Applying a similar correction factor to the estimated interaction effect (after one year) yields an interaction effect of 0.13 standard deviations, hence the difference in the impact of textbooks between a student one standard deviation below and a student one standard deviation above the mean academic achievement would be 0.26 standard deviations.

This attenuation bias can potentially be corrected by instrumenting the pre-test scores with scores in other subjects. Instrumental variable estimates are consistent, however, only if pre-test scores are correlated across subjects and, conditional on the pre-test score in subject k, a student's pre-test score in another subject has no effect on his or her post-test score in subject k. The second assumption could be challenged. For example, skills in English may help students learn math and science, and math skills maybe useful in science. Both possibilities violate the exclusion restriction required for consistent estimates of differential program effects in math and science. Yet it is unlikely that math and science skills help in English, which implies that estimates of differential program effects in English should be consistent.

The point estimate of the textbook effect using instrumental variables is 0.14 standard deviations in year 1 and 0.13 standard deviations in the year 2, both of which are significant at the 1% level (Table 14). The estimated effects for years 3 and 4 are smaller and not statistically significant (not shown in Table 14). Given potential problems with instrumenting for math scores with English scores, we ran separate regressions in which math scores were instrumented only with science scores. The results were very similar, both in magnitude and significance.

The estimates in Tables 13 and 14 constrain the interaction between pre-test scores and the program to be linear. They also constrain the school random effects to be the same for students of different abilities, which rules out the possibility that some schools are particularly good for high-ability students but bad for low-ability students. Table 15 relaxes these assumptions by dividing the sample into quintiles, according to average pre-test scores, and re-estimating the level regressions for years 1 and 2 in Table 9 for each quintile. This allows for separate treatment effects, and different school random effects, for each quintile.

Aggregating across all subject-grade combinations, the estimated effects of textbooks on test scores in year one for the five quintiles, from lowest to highest, are -0.05, -0.02, 0.03, 0.14 and 0.22 standard deviations. These parameters are statistically insignificant for the first three quintiles, but those for the fourth and fifth are significant at the 10% and 5% levels, respectively. Since the pre-tests were noisy, some of the students in the top quintile of pre-test scores were not necessarily in the top quintile of initial learning. Thus if the effect of textbooks does depend on initial achievement, the true effect in the top quintile of initial learning will be greater than 0.22 standard deviations. The pattern of effects across quintiles in year two is quite similar, except that the effects for the top two quintiles are less precisely estimated and statistically insignificant. The pattern in the third year (not shown in Table 15) is even weaker, and in the fourth year (also not shown) disappears altogether, which is consistent with the results in Tables 13 and 14.

Textbooks may be more beneficial for students with strong initial scores either because pupils with a stronger academic background are better able to use the textbooks or because better schools use textbooks more effectively. We do not find support for an alternative hypothesis that better motivated children are more likely to take the textbooks home to work with them, and also more likely to do well, creating a positive interaction.

The government textbooks provided through the program may have been too difficult for many students. For example, all the textbooks are written in English, which is the third language for the vast majority of students. This alone may make if difficult for many students to use the textbooks effectively. Indeed, there is evidence that the median students in the lower grades of these schools had difficulty reading English. A random sample of 50 of the 100 schools were visited in the middle of year 4, and in grades 3-8 the median student (by class rank) was asked to read the English textbook provided by the program. In grade 3, only 16% of the median students were able to read the grade 3 English textbook, and only 28% of the median students in grade 4 could read the English textbook for that grade. This is less common for upper grade students -- the analogous figure for grade 5 is 67%, and for grades 6-8 it is over 90%. Yet many students, especially in grades 3 and 4, are likely to have had difficulty using textbooks effectively.

Theoretically, differences in student motivation could also contribute to the interaction effect. If more motivated students took the textbooks home and utilized them effectively, and these same students had higher pre-test scores, also due to better motivation, we would see a positive interaction between textbooks and pre-test scores. In fact, we find that children in the program schools who actually took the textbooks home had higher average pre-test scores (0.21) than did students in those school who did not take them home (-0.03), a difference that it statistically significant at the 1% level.

However, this differential in the propensity to take books home appears to explain at most only a small proportion of the interaction effect. To check this, the year 1 interaction effect over all subjects in Table 13 (0.057) was estimated separately for grades 3-5 (schools were asked by ICS not to allow students in these grades to bring textbooks home) and grades 6-8 (no such request was made for students in these grades). Statistically significant interaction effects were

found in both estimates, and that for the lower grades (0.051) was only slightly smaller than the estimate for the upper grades (0.072).

To conclude, one of the two possible explanations for why stronger students benefited the most from the textbooks was supported by the evidence: many students in the lower grades appeared to be unable to read the English textbooks provided by the program. The second possibility, that better students benefit from being able to take the textbooks home and they are more motivated to do so, received some support but probably plays only a secondary role.

There are several possible reasons why the program effect did not persist past the first two years. First, textbook availability in the comparison schools increased in years three and four (Table 4). For example, in the first year the book to pupil ratio was 0.74 in treatment schools, compared with 0.19 in comparison schools (for subjects receiving textbooks), but by year 3 it is only 0.61 in treatment schools and 0.25 in comparison schools. This partly reflects provision of textbooks to a few schools in mid-1997 (year 2), and perhaps a general trend among schools in Busia to purchase more textbooks (new editions of some textbooks appeared in 1997). Given the classroom setup and the fact that students share benches, two or three students can share a book with only modest limits on their ability to use it. Given this, the introduction of textbooks into the comparison schools would not have to be that dramatic in order to see convergence in the effects. This is particularly true given that the effect of textbooks seems to be largely due to higher-ability students, who we might reasonably expect to be given access to any new texts provided.

Second, we see convergence in teacher attendance and textbook usage between years 2 and 3 (Table 6). If increased textbook usage in class is contributing to the differences in test scores, this could account for at least part of the convergence. This is combined with a general

decrease in textbook use in the textbook schools between years 2 and 3. Sixty-two percent report using textbooks in year 2 (in the textbook schools) versus only 45% in year 3. This may reflect the issuance of new textbooks during this period – teachers may have been reluctant to continue to have students use the old books, despite the fact that they were quite similar.

A third explanation for the temporary effect is the decreasing ability of pre-test scores to measure ability or motivation. All interaction effects are based on the January 1996 pre-tests, which may decline over time in their accuracy in measuring true ability or motivation. Children's academic ability may vary over time as they mature at different rates and as new material is learned. Similarly, motivation can also change over time. This hypothesis is also consistent with the steady decline over time in the coefficients on the pre-test score in Tables 8 and 9.

A fourth possible explanation is that the sample size becomes smaller each year as children drop out or complete their primary education, which reduces the precision of the estimated impacts of the program. A fifth and final possibility is that the crowding out tended to take place primarily in capital investments. If textbooks are something that depreciates relatively quickly, relative to new construction, for example, then one may not see the effect of crowding out on test scores after a single year, but the cumulative impact after several years may become more substantial.

To summarize Section VI, there is no evidence that the program increased test scores for the average student. Our three estimators for the impact on average student achievement allow us to reject average effects greater than 0.20, 0.13 and 0.07 standard deviations for year one, and data from later years provide no evidence that the impact of textbooks on test scores accumulates over time. Yet we do find evidence of increases in test scores for strong students during the first two years of the program, although this effect does not extend into the program's third and fourth

years. These findings are reinforced by those in a related paper by Glewwe, Iyer and Kremer (2004). That paper examined the impact of providing grants to the second group of schools in 1997; those grants could be used to purchase a wide variety of educational materials, and for construction, but most schools opted to purchase textbooks. The results are strikingly similar: there is no effect on the average student, but students who were already doing well had significantly higher test scores.

VII. Summary and Policy Implications

Provision of textbooks to schools where few students currently have them would seem to most observers to be an obvious way to raise students' educational performance in developing countries. To our knowledge, this policy recommendation is almost universally accepted, even by those who harbor general doubts about the effectiveness of increases in school spending. Yet our results show that provision of textbooks in Kenya increased scores only for strong students. This probably reflects the greater ability of stronger students take advantage of the texts.

The results here are at odds with previous retrospective studies, which typically find that provision of textbooks had positive effects on student performance, as well as with the retrospective analysis in Section III. One possible interpretation of the retrospective results is that parents are more likely to purchase textbooks for their children if they think they that their children are high-skill students, which leads to a situation where textbooks reinforce already existing skill differences and the recent estimates of the treatment effect to be biased upward.[17]

[17] It is worth contrasting our results with those of Jamison et al [1988] in Nicaragua who found much more positive impacts. The students in Nicaragua received workbooks, in which they could write the answers to mathematics problems. Workbooks allow for more active learning, but are much more expensive than textbooks. One workbook is needed per student, whereas one textbook may serve for two or three students in rural Kenyan schools. Moreover,

This analysis may also be useful in interpreting the impact of a program which provided textbooks and paid for the school uniforms that children in Kenya are required to purchase. The program led to a 40% increase in enrollment. The joint impact of textbook provision and a 40% increase in enrollment on test scores was small, but based on the earlier study it was impossible to determine whether textbooks had a large positive impact on test scores which was offset by a large negative impact of increased enrollment, or whether the effects of both textbooks and class size were small. This study suggests that it is more likely that both effects were small.

There is no reason to believe that textbooks inherently help exclusively stronger students. In fact, some studies suggest that inputs often disproportionately help the weaker students. For example, Krueger (1999) found that, among K-3 students in Tennessee, increases in test scores from smaller class size are highest for weaker students. Case and Deaton (1999) find that smaller class size has a significant effect on test scores among Black students age 10-18 in South Africa, but not for Whites, whose average scores are much higher. The effect of textbooks on strong students may reflect the nature of Kenya's educational system, but since the system is similar to those in other poor developing countries, the results are relevant for many countries.

This raises the question of what policies might be appropriate and feasible. Lowering the level of rigor in the existing curriculum may not be politically feasible, given the political power of more advantaged groups and the possibility that they could simply exit the public education system altogether. One potential way out of the dilemma is to allow different schools or different programs within schools, to progress through the curriculum at different rates. Thus, for

---

a workbook can only be used once, whereas the same textbook can be used for three or four years. Thus the annual cost of workbooks per student may be six to twelve times that of textbooks. Another difference is that Nicaraguan teachers received three hours of training in the use of workbooks whereas the Kenyan teachers received none. Of course, the particular textbooks given and the schools and parental backgrounds also differ between the studies. In

example, some schools could seek to cover the primary school curriculum, which is currently supposed to be covered in eight years, in seven, while others might seek to cover it in nine or even ten years. Since all students would take the same curriculum, students would not be separated at entrance into primary school or at some other early age into those destined for academic careers and those destined for second-rate education (or at least not any more so than is currently the case). Students in schools that completed the KCPE after nine or ten years would have the same opportunity to enter secondary school as those in schools that complete the KCPE in seven years. In Singapore, more able students are placed in an “express” track, and after four years take the national exam (GCE-‘O’), which allows them to apply to junior colleges. The rest are placed in the “normal” track, which leads to an intermediate exam (GCE-‘N’) at the end of four years. Those who are competent go on to take the GCE 'O' examination at the end of the fifth year (Ministry of Education of Singapore, Web Page).

In practice, the Kenya education system already involves students completing the same material at very different paces, in a highly inefficient way. Even though officially everyone in Kenya faces an eight-year curricula, in practice, many students are already on the ten-year plan. Repetition rates are in Kenya are over 20% (see Table 8) However, it is inefficient to cover the same material at two different times, a year apart, first racing through it at a speed that doesn't allow students to comprehend, and moving on to more difficult topics that require having mastered the original, and then returning the next year to material previously covered. For example, a student who hasn't successfully learned multiplication tables may be taught how to multiply two-digit numbers rather than spending more time on learning multiplication tables before going on to learning the two-digit numbers..

---

the Kenyan case the textbooks used were written in a foreign language and were difficult for many students to read.

Another way in which Kenyan schools already effectively deliver heterogeneous education is by teachers not covering the entire curriculum in a class year in some schools. It's quite routine for teachers to only progress partly through the curriculum for that year in a given year. However, structuring education this way is inefficient, since the teacher in the next grade is likely not to pick up halfway through the curriculum of the previous grade, but rather to start with the curriculum for their grade (i.e. even if the sixth-grade teacher only gets halfway through the sixth-grade textbook and the sixth-grade curriculum, the seventh-grade teacher is likely to start at the beginning of the seventh-grade curriculum—not to teach sixth-grade material). This tendency is of course reinforced by the practice of giving exams that cover the assigned curriculum for that year. A seventh-grade teacher who decides to help students complete the sixth-grade curriculum is likely to wind up with students that score very low on the grade 7 district exams. The teacher is not likely to be held in high esteem by her headmaster, her colleagues, or the community in which she teaches. In contrast, if the teacher covers the seventh grade curriculum, at least some students will do well.

One piece of evidence that the Kenyan system is not unique in being oriented away from the needs of the typical student comes from Banerjee and Duflo's work in India on remedial education, which suggests that paying close attention to lower-performing students can have big effects for them (Banerjee, Cole, Duflo and Linden, 2003).

Table 1: Cross-Sectional Retrospective Estimates of Impact of Textbooks on Test Scores

	English	*Math*	*Science*	*All*
Student-owned books	0.178***	0.087***	0.054	0.116***
	(0.024)	(0.026)	(0.047)	(0.016)
School-owned books	-0.010	-0.354***	-0.554	-0.190**
	(0.170)	(0.145)	(0.442)	(0.095)

* significant at 10%, ** significant at 5%, *** significant at 1%.

Notes: Asymptotic standard errors are in parentheses. All regression included school random effects, a constant term and dummy variables for grade. Other variables included in each regression are parental education (for students in grades 6-8, for whom data are available), wealth (as measured by land owned), parental participation in school, teacher education, and teacher training. The test scores used are the October 1996 ICS tests for grades 3 and 4, the October 1996 district mock tests for grades 5-7 and the November 1996 KCPE tests for grade 8.

Table 2: Longitudinal Retrospective Estimates on Impact of Jomo Kenyatta Foundation Textbooks on Test Scores

Grade	*1994*	*1995*
6	-0.157	-0.091
	(.171)	(.266)
7	0.497**	0.641**
	(.252)	(.291)
8	0.020	.676***
	(.172)	(.204)

Notes: Standard errors are in parentheses; * significant at 10%, ** significant at 5%, *** significant at 1%.
Each cell represents a regression of normalized test score in 1994 or 1995 minus normalized test score in 1993 on geographic dummy variables and a dummy variable for whether the school received textbooks (coefficient on that dummy is reported).

Scores are normalized across students, not across schools: although the data are school mean scores they have been standardized in terms of the distribution across students using a sample of students from the SAP schools in each of the years.
Sample sizes ranged from 255 to 274.

Table 3: Differences in Normalized Pre-Test Scores Between Textbook Schools and 25-School Comparison Group[a]

Subject	*Grade*	*Difference between textbook and comparison schools*	*Standard error*	*Number of observations*
English	Grades w/ texts[a] (3-7)	0.046	0.105	8,516
	All grades[a] (3-8)	0.033	0.101	9,332
Math	Grades w/ texts[a] (3, 5 and 7)	0.056	0.090	5,069
	All grades[a] (3-8)	0.054	0.085	9,302
Science	Grades w/ texts[a] (8)	0.173	0.105	816
	All grades[a] (3-8)	-0.017	0.088	9,276
All subjects combined[b]	Grades w/ texts	0.061	0.091	14,401
	All grades	0.023	0.087	27,910

Notes: Each row represents a regression of January 1996 pre-test scores on a constant and a dummy variable for being in a textbook school, with school random effects. The sample consists of all students from the 25 textbook schools and the 25-school comparison group who took the January 1996 pre-test.
[a] These rows stack across different grades and includes dummy variables for each grade.
[b] The row labeled "All grades" stacks across subjects and grades and has dummy variables for each grade/subject combination. The row "Grades w/ texts" excludes grade/subject combinations that did not receive textbooks.

Table 4: Availability of Textbooks
(book per pupil ratio)

		School-owned books		**Privately-owned books**		**School + Private books**	
Program Year	*Subject/Grade where textbooks provided?*	*Textbook schools*	*Comparison schools*	*Textbook schools*	*Comparison schools*	*Textbook schools*	*Comparison schools*
1	Yes	0.65	0.04	0.09	0.15	0.74	0.19
	No	0.02	0.03	0.05	0.07	0.08	0.09
2	Yes	0.55	0.04	0.09	0.17	0.64	0.21
	No	0.04	0.03	0.08	0.08	0.12	0.12
3	Yes	0.52	0.11	0.08	0.14	0.61	0.25
	No	0.11	0.09	0.09	0.09	0.20	0.19
4	Yes	0.43	0.10	0.05	0.11	0.48	0.21
	No	0.10	0.08	0.05	0.06	0.17	0.14

Notes: Textbook availability is calculated using school questionnaire data collected at the start of each school year, and data on privately-owned textbooks from a pupil questionnaire given to pupils in grades 6-8 (privately-owned textbooks for pupils in grades 3-5 are from the school questionnaire).
Results for 1996, 1997 and 1998 aggregate over grades 3-8 for and over three subjects: English, math and science.

For 1999, results are for grade 6 to 8 only in Math and English. Comparison schools varied by year: the 75-school

comparison group in 1996, the 50-school group in 1997 and the 25-school group in 1998 and 1999.

Table 5: Student Reporting on Availability of School-Owned Textbooks (Grades 6-8 only)

		School issued you a textbook to use in class?		**School allowed you to take textbook home?**	
Year	*Type of subject/grade combination:*	*Textbook schools (%)*	*Comparison schools (%)*	*Textbook schools (%)*	*Comparison schools (%)*
2	Textbooks provided	62.4	7.7	52.8	2.5
2	Textbooks not provided	8.6	7.1	5.4	1.9
3	Textbooks provided	72.0	28.3	63.5	9.4
3	Textbooks not provided	23.4	11.7	17.4	6.4

Note: These figures are averages over groups of grade/subject combinations, disaggregated according to whether the combination received textbooks from ICS. In both years, "textbooks provided" refers to English and math in grades 6 and 7 and science in grade 8, while "textbooks not provided" refers to science in grades 6 and 7 and English and math in grade 8. This information is available only from the 1997 and 1998 student questionnaires, which were administered to students in grades 6-8 in those years. Similar data were collected in 1996, but for the textbook schools all the interviews were conducted in January and February of 1996, which is before or during the time that the textbooks were being distributed (February 1996).

Table 6: Impact of Program on Observed Textbook Usage and Pedagogy

	Year 2		Year 3	
	Textbook schools (%)	*Comparison schools (%)*	*Textbook schools (%)*	*Comparison schools (%)*
Teacher absent	24.1**	37.6	32.3	31.0
Pupils used textbook in class:				
Unconditional	62.2**	45.8	45.6	37.4
Conditional on teacher present	78.9*	67.8	64.1**	51.7
Days doing homework during last 7 days (pupils in grades 6-8)	4.7**	4.1	4.2**	3.8
Distribution of pupils by sharing of textbooks, conditional on using them:				
Own textbook	8.4	3.9	8.1	2.6
2 Students share	35.8	16.3	29.1	18.7
3 Students share	30.1	28.8	37.7	29.0
4 Students share	16.5	21.5	12.0	19.6
5 Students share	5.3	17.5	8.1	17.7
No textbook	4.1	12.0	5.0	12.3
Implied textbook/student ratio	0.41	0.30	0.40	0.29

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Figures are averages over grade/subject combinations that had received textbooks by 1997 (grades 3-7 for English and math and grade 8 for science). The comparison schools were the 50-school comparison group in 1997 and the 25-school group in 1998. Statistical significance for teacher absence and teacher use of textbooks was obtained from random effects probit estimates.

Table 7: Assistance Received from ICS and from Other Sources in 1996

	Textbook schools[a] *(U.S. dollars)*	*Comparison schools*[a] *(U.S. dollars)*	*Difference*[a] *(U.S. dollars)*
All schools			
ICS assistance:			
Mean	485	--	--
	(46)		
Standard deviation	231	--	--
Non-ICS assistance:			
Mean	267	456	-189
	(94)	(154)	(272)
Standard deviation	470	1330	--
Number of schools	25	75	--
Large schools			
Non-ICS assistance:			
Mean	421	291	130
	(153)	(152)	(265)
Standard deviation	574	923	--
Number of schools	14	37	--
Small schools			
Non-ICS assistance:			
Mean	71	606	-535
	(50)	(271)	(502)
Standard deviation	165	1650	--
Number of schools	11	37	--

[a] Standard errors shown in parentheses.

Table 8: Promotion, Repetition, Dropout and Attendance Rates

Lower grades (Children who would still be in primary school if grade progression stayed on schedule)

	Year 2		***Year 3***		***Year 4***	
	Textbook schools	*Comparison schools*	*Textbook schools*	*Comparison schools*	*Textbook schools*	*Comparison schools*
Stayed, promoted	.53	.53	.31	.30	.20	.19
Stayed, repeated 1 year	.21	.21	.29	.27	.29	.29
Stayed, repeated 2+ years	-	-	.03	.02	.09	.07
Dropped out	.17	.17	.28	.28	.34	.35
Transferred out	.08	.09	.08**	.11	.08	.09
Number of students	5009	4838	3941	3778	2895	2803

Higher grades (Children who would have finished primary school if grade progression were on schedule)

	Year 2		***Year 3***		***Year 4***	
	Textbook schools	*Comparison schools*	*Textbook schools*	*Comparison schools*	*Textbook schools*	*Comparison schools*
Finished primary, no secondary	.32***	.41	.26	.30	.24	.23
Entered secondary	.43**	.38	.28	.29	.25	.24
Stayed, repeated 1 year	.16	.14	.15***	.11	.08	.09
Stayed, repeated 2+ years	-	-	.04	.03	.08	.06
Dropped out	.01	.03	.22	.23	.32	.35
Transferred out	.06	.04	.04	.04	.03	.04
Number of students	447	440	1,182	1,179	1,867	1,770

Attendance for all grades combined

	Year 2		***Year 3***		***Year 4***	
	Textbook schools	*Comparison schools*	*Textbook schools*	*Comparison schools*	*Textbook schools*	*Comparison schools*
Attendance Rate	83.9	81.5	79.7	77.8	83.5	84.5
Number of students	4959	9707	3558	3315	2652	2516

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
For promotion, repetition and dropping out, comparison schools are always the 25-school comparison group. Lower and upper grades refer to grade of pupils in 1996. Lower (upper) grades are 3-7 (8) for 1997, 3-6 (7 and 8) for 1998 and 3-5 (6-8) for 1999. Tests for statistical significance are based on probits with school random effects For attendance, year 2 figures compare the 25 textbook schools to the 50-school comparison group, with statistical significance based on probit regressions with school random effects. Years3 and 4 results compare the 25 textbook schools to the 25-school comparison group, and tests for statistical significance are GLS estimates of the number of days attended (0 to 4) with school random effects. For all three years, attendance is conditional on enrollment.

Table 8: Attendance Rates in Years 2, 3, and 4

		Year 2 (one visit)		Year 3 (four visits)		Year 4 (three visits)	
		Textbook schools (%)	*Comparison schools (%)*	*Textbook schools (%)*	*Comparison schools (%)*	*Textbook schools (%)*	*Comparison schools (%)*
Grade	3	81.9	80.9	79.9	79.0	71.9	75.3
	4	81.5**	79.9	77.6	74.7	77.8*	83.1
	5	84.3	81.2	77.8	75.1	81.0	80.2
	6	84.2	81.7	78.1	76.5	83.6	84.5
	7	86.0	82.0	83.6	80.1	85.7	86.4
	8	89.3	85.9	85.7*	90.6	95.0	93.1
All Grades		83.9	81.5	79.7	77.8	83.5	84.5

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Year 2 figures compare the 25 textbook schools with the 50-school comparison group. Tests for statistical significance in year 2 are based on probit regressions with school random effects. Years 3 and 4 figures compare the 25 textbook schools with the 25-school comparison group. Tests for statistical significance in years three and four are GLS estimates of the number of days attended (0 to 4) with school random effects. For all three years, attendance is conditional on enrollment.

Table 9: Impact of Textbook Program on Normalized Test Scores

Program year	*Subject*	*Grades*	*Years exposed to textbooks*	*Coefficients on textbook schools (β)*	*Standard error*	*Number of observations*
1	English	3-7	1	-0.001	0.104	14,253
	Math	3,5,7	1	0.055	0.082	8,408
	Science	8	1	0.033	0.118	1,471
	All subjects	3-8	1	0.022	0.086	24,132
2	English	4-7	2	0.063	0.112	8,098
	English	3,8	1	-0.011	0.118	3,342
	English	3-8	1 or 2	0.041	0.104	11,440
	Math	4,6	2	-0.057	0.103	4,389
	Math	3,5,7,8	1	0.059	0.084	6,918
	Math	3-8	1 or 2	0.017	0.088	11,307
	Science	8	1	0.185**	0.087	1,409
	All subjects	3,5,7,8	1	0.047	0.082	11,669
	All subjects	4-7	2	0.023	0.105	12,487
	All subjects	3-8	1 or 2	0.032	0.091	24,156
3	English	5-7	3	-0.021	0.121	3,752
	English	4,8	2	-0.099	0.105	1,969
	English	4-8	2 or 3	-0.052	0.109	5,721
	Math	5,7	3	-0.022	0.108	2,486
	Math	4,6	2	-0.023	0.087	2,418
	Math	4-7	2 or 3	-0.024	0.092	4,904
	Science	8	1	-0.086	0.099	817
	All subjects	4,6,8	2	-0.053	0.089	4,387
	All subjects	5-7	3	-0.023	0.112	6,238
	All subjects	4-8	2 or 3	-0.042	0.098	10,625
4	English	6,7	4	-0.047	0.154	2,146
	English	5,8	3	-0.058	0.136	1,769
	English	5-8	3 or 4	-0.056	0.138	3,915
	Math	6	4	-0.121	0.134	1,183
	Math	5,7,8	3	-0.048	0.100	2,732
	Math	5-8	3 or 4	-0.074	0.105	3,915
	Both subjects	5,7,8	3	-0.051	0.103	4,501
	Both subjects	6,7	4	-0.085	0.145	3,329
	Both subjects	5-8	3 or 4	-0.065	0.114	7,830

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents a random effects regression of October/November test scores from the relevant year on a constant and a dummy variable for being in a textbook school, and sex and region dummy variables. The sample consists of all children who were enrolled in January 1996 and took the relevant October/November test. Years 1, 2, 3, 4 refer to 1996, 1997, 1998, and 1999, respectively.

Table 10: Impact of Textbook Program on Normalized Test Scores Minus Pre-Test Scores

Program year	*Years exposedto textbooks*	*Coefficients on textbook schools (β)*	*Standard error*	*Number of observations*
1	1	0.019	0.053	11,321
2	1	0.072	0.057	4,678
	2	-0.039	0.070	7,377
	1 or 2	0.009	0.051	12,055
3	2	0.037	0.080	1,874
	3	-0.053	0.078	5,533
	2 or 3	-0.033	0.066	7,407
4	3	-0.019	0.102	2,231
	4	-0.188	0.127	2,908
	3 or 4	-0.107	0.092	5,139

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents a random effects regression of October/November test scores from the relevant year minus pre-test scores on a constant and a dummy variable for being in a textbook school, and sex and region dummy variables. The sample includes all children enrolled in January 1996 who took both the January 1996 pre-test and the relevant October/November test. All results are aggregated over all grade-subject combinations that received textbooks.

Table 11: Selection and Attrition During the First Year of the Program

	Textbook (%)	*Comparison (%)*	*Difference[a] (%)*
Drop outs and transfers from 1995 to January 1996 (20 schools)			
Drop outs	5.3	6.0	-0.7
Transfers out	5.2	3.6	1.6
Composition of students, January 1996 (50 schools)			
Repeaters[a]	21.9	26.0	-4.1***
Transfers in	11.2	10.3	0.9
Students in January 1996 not tested in October of (50 schools):[b]			
First year	26.0	26.3	-0.3
Second year	31.0	33.3	-2.3***
Third year	38.6	39.9	-1.2
Fourth year	45.2	47.9	-2.7

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
These inferences are based on probit regressions with school random effects. The explanatory variables include a constant term, a dummy variable for textbook schools, a dummy variable for sex and dummy variables for each grade. The regressions for 50 schools add dummy variables for the seven geographic divisions in Busia.
[a] The percentage of repeaters is slightly underestimated for both types of schools because there is no information on repetition for nearly one half of students who transferred in (about 6% of all students).
[b] The first year results on whether students were tested include all students. The second year results exclude students who were in grade 8 in the first year (1996), since most were no longer in school and could not be tested. For the same reasons, the third year results exclude children in grades 7 and 8 in the first year and the fourth year results exclude children in grades 6, 7 and 8 in 1996.

Table 12: Relative Normalized Test Scores After One Year by Whether Subject-Grade Combinations Received Textbooks

Grade/ subject	*Received a textbook*	*Textbook school*	*Number of observations*
3 (All subjects)	-0.085	0.044	11,430
	(0.108)	(0.132)	
4 (All subjects)	0.068	0.025	10,545
	(0.089)	(0.127)	
5 (All subjects)	0.116	-0.039	6,608
	(0.094)	(0.114)	
6 (All subjects)	-0.176*	0.124	7,004
	(0.095)	(0.098)	
7 (All subjects)	0.026	0.018	7,114
	(0.092)	(0.115)	
8 (All subjects)	0.044	0.003	4,415
	(0.091)	(0.106)	
English (All grades)	0.045	-0.046	15,726
	(0.127)	(0.146)	
Math(All grades)	0.037	0.022	15,732
	(0.075)	(0.084)	
Science (All Grades)	-0.021	0.067	15,658
	(0.121)	(0.084)	
All subjects, all grades	-0.009	0.036	47,116
	(0.040)	(0.083)	

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents a random effects regression of October 1996 test scores on dummy variables indicating: a) whether textbooks were provided for a specific grade/subject combination; b) whether school received textbooks; and c) region and sex. The sample includes all children enrolled in the 25 textbook schools and the 75-school comparison group in January 1996 who took the October 1996 post-tests. Standard errors are shown in parentheses.

Table 13: Normalized Test Scores as a Function of Treatment and Pre-Test Score

				Textbook school			*Textbook *pre-test*		
Year	*Subject*	*Grade*	*Years exposed*	*Coeff.*	*Std. err.*	*Pre-test scores*	*Coeff.*	*Std. err.*	*Number of observations*
1	English	3-7	1	0.047	0.083	0.404	0.069***	0.022	6,710
	Math	3,5,7	1	0.043	0.076	0.440	0.077**	0.032	3,954
	Science	8	1	0.178*	0.112	0.534	-0.076	0.059	678
	All subjects	-	1	0.060	0.061	0.429	0.057***	0.018	11,342
2	English	4-7	2	0.037	0.093	0.318	0.104***	0.026	4,818
		8	1	0.049	0.084	0.419	-0.021	0.052	851
		4-8	1 or 2	0.039	0.080	0.335	0.081***	0.023	5,669
	Math	4,6	2	-0.109	0.092	0.339	-0.013	0.038	2,575
		5,7,8	1	0.107	0.077	0.324	0.061*	0.032	2,991
		4-8	1 or 2	0.014	0.074	0.352	0.029	0.025	5,566
	Science	8	1	0.207***	0.079	0.404	-0.000	0.050	842
	All subjects	-	1	0.106*	0.057	0.356	0.034	0.025	4,684
		-	2	-0.014	0.083	0.345	0.064***	0.022	7,393
		-	1 or 2	0.035	0.067	0.348	0.049***	0.017	12,077
3	English	5-7	3	-0.022	0.098	0.287	0.023	0.029	3,344
		5-8	2 or 3	-0.036	0.086	0.300	0.014	0.026	4,085
	Math	5,7	3	0.054	0.087	0.384	-0.033	0.036	2,207
		5-7	2 or 3	0.009	0.076	0.286	0.007	0.030	3,342
	All subjects	-	2	-0.010	0.064	0.211	0.053	0.037	1,876
		-	3	-0.016	0.084	0.325	0.003	0.023	5,551
		-	2 or 3	-0.020	0.073	0.297	0.006	0.020	7,427
4	English	6,7	4	-0.062	0.145	0.098	0.054	0.042	2,146
		8	3	-0.023	0.120	0.111	0.032	0.082	765
		6-8	3 or 4	-0.041	0.127	0.097	0.055	0.038	2,911
	Math	6	4	-0.114	0.125	0.250	-0.063	0.059	1,183
		7,8	3	-0.036	0.112	0.176	-0.053	0.049	1,728
		6-8	3 or 4	-0.072	0.101	0.199	-0.053	0.038	2,911
	All subjects	-	3	-0.021	0.096	0.152	-0.029	0.043	2,493
		-	4	-0.092	0.133	0.147	0.019	0.035	3,329
		-	3 or 4	-0.059	0.102	0.148	-0.001	0.027	5,822

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents a random effects regression of October/November test scores from the relevant year on dummy variables for being in a textbook school, sex and region, the average of the three January 1996 pretest scores, and an interaction term between the textbook school dummy and the average pretest score. The sample consists of all children who have enrolled in January 1996 and took both the January 1996 pre-test and the relevant October/November test.

Table 14: Normalized Test Scores as a Function of Treatment and Pre-Test Score (Instrumental Variable Estimates)

				Treatment			*Treatment*pre-test*		
Year	*Subject*	*Grade*	*Years exposed*	*Coeff.*	*Std. err.*	*Pre-test scores*	*Coeff.*	*Std. err.*	*Number of observations*
1	English	3-7	1	-0.033	0.077	0.890	0.108	0.074	6,631
	Math	3,5,7	1	-0.047	0.080	0.701	0.235***	0.079	3,898
	Science	8	1	0.048	0.598	1.361	-0.111	0.291	678
	All subjects	-	1	-0.035	0.066	0.839	0.144***	0.055	11,211
2	English	4-7	2	0.032	0.091	0.797	0.139	0.086	4,772
		8	1	-0.023	0.112	0.847	-0.136	0.137	847
		4-8	1 or 2	0.021	0.086	0.820	0.065	0.076	5,619
	Math	4,6	2	-0.145	0.093	0.719	-0.069	0.090	2,542
		5,7,8	1	0.062	0.068	0.682	0.408***	0.133	2,974
		4-8	1 or 2	-0.032	0.062	0.663	0.147**	0.075	5,516
	Science	8	1	0.225***	0.085	0.647	0.061	0.114	838
	All subjects	-	1	0.067	0.044	0.677	0.216***	0.077	4,659
		-	2	-0.036	0.078	0.745	0.089	0.066	7,141
		-	1 or 2	0.010	0.056	0.718	0.131***	0.051	11,973

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents a random effects regression of October/November test scores from the relevant year on dummy variables for being in a textbook school, sex and region, the January 1996 pretest score (instrumented by the score on the two other tests), and an interaction term between the textbook school dummy and the pretest score. The sample consists of all children who have enrolled in January 1996 and took both the January 1996 pre-test and the relevant October/November test.

Table 15: Normalized Test Scores by Quintile of Pre-Test Scores

Year	*Years exposed*	*Quintile 1*	*Quintile 2*	*Quintile 3*	*Quintile 4*	*Quintile 5*
1	1	-0.049	-0.021	0.032	0.142*	0.218**
		(0.064)	(0.069)	(0.073)	(0.079)	(0.096)
2	1	0.056	0.085	0.018	0.182**	0.207*
		(0.073)	(0.074)	(0.090)	(0.080)	(0.122)
	2	-0.080	-0.096	-0.090	0.022	0.173
		(0.081)	(0.094)	(0.103)	(0.100)	(0.133)
	1 or 2	-0.029	-0.028	-0.033	0.099	0.188
		(0.067)	(0.070)	(0.088)	(0.082)	(0.116)

Notes: * significant at 10%, ** significant at 5%, *** significant at 1%
Each row represents five random effects regressions, one for each quintile (based on January 1996 pre-test scores) of post-test scores on a dummy variable indicating whether a child is in a textbook school and on dummy variables for region and sex. The sample consists of all children enrolled in January 1996 who took both the 1996 pre-test and the relevant post-test score. All results are aggregated over all grad subject combinations that received textbooks.

Table A.1: Tests Administered from 1996 to 1999

Test	*Date administrated*	*Administered by*[1]	*Schools*	*Grades*
1996 Pre-test	January, 1996	ICS	50 (Groups 1 and 4)	3-8
1996 Post-test	October, 1996	ICS	100	3, 4
1996 Mock test	October, 1996[2]	DEO	100	5-7
1996 KCPE	November, 1996	DEO	100	8
1997 Post-test	October, 1997	ICS	100	3-7
1997 KCPE	November, 1997	DEO	100	8
1998 Post-test	October, 1998	ICS	100	3-8
1998 Mock test	October, 1998[2]	DEO	68[3]	4-7
1998 KCPE	November, 1998	DEO	100	8
1999 Post-test	October, 1999	ICS	100	3-8
1999 Mock test	October, 1999[2]	DEO	100	4-7
1999 KCPE	November, 1999	DEO	100	8

Notes:
1. ICS = Internationaal Christelijk Steunfonds; DEO = District Education Office of Kenya Ministry of Education.
2. Mock tests were also administered in grade 8 in July of 1996, 1998 and 1999. These are not used in our analysis.
These mock tests were not administered in Teso district, where 32 of the schools are located.

Table A2: Understanding of the English Textbook

Grade	*Can read the book (%)*	*Can answer questions, in English, about the passage (%)*	*If unable to answer question in English, can answer when asked in Kiswahili (%)*	*Can answer written questions, in English, from the book (%)*
3	16	10	51	37
4	28	10	61	12
5	67	49	51	47
6	92	86	82	39
7	96	77	22	56
8	100	99	-	92

Note: Each student was the median student in their class at school. The data on answering questions in English include all students. Those unable to read the passage had it read to them in English.

Table A3: Differences in Normalized Pre-Test Scores Between Textbook Schools and 25-School Comparison Group[a]

Each row represents a regression of January 1996 pre-test scores on a constant and a dummy variable for being in a textbook school, with school random effects. The sample consists of all students from the 25 textbook schools and the 25-school comparison group who took the January 1996 pre-test.

Subject	*Grade*	*Received textbooks?*	*Difference between textbook and comparison schools*	*Standard error*	*Number of observations*
English	3	Yes	-0.032	0.147	2,054
	4	Yes	0.027	0.099	1,945
	5	Yes	0.090	0.146	1,625
	6	Yes	-0.014	0.125	1,480
	7	Yes	0.186	0.126	1,412
	8	No	-0.041	0.132	816
	Grades w/texts[a]		0.046	0.105	8,516
	All grades[a]		0.033	0.101	9,332
Math	3	Yes	0.175	0.136	2,033
	4	No	0.027	0.120	1,940
	5	Yes	0.014	0.135	1,625
	6	No	0.059	0.137	1,477
	7	Yes	-0.020	0.091	1,411
	8	No	0.098	0.116	816
	Grade w/texts[a]		0.056	0.090	5,069
	All grades[a]		0.054	0.085	9,302
Science	3	No	-0.163	0.161	2,026
	4	No	-0.087	0.116	1,932
	5	No	0.017	0.134	1,622
	6	No	0.036	0.096	1,473
	7	No	-0.007	0.142	1,407
	8	Yes	0.173	0.105	816
	All grades[a]		-0.017	0.088	9,276
All subjects combined[b]	3		-0.005	0.122	6,113
	4		-0.022	0.098	5,817
	5		0.048	0.128	4,872
	6		0.028	0.103	4,430
	7		0.048	0.102	4,230
	8		0.071	0.083	2,448
	All grades		0.023	0.087	27,910
	All grades w/texts		0.061	0.091	14,401

[a] This row stacks across different grades and includes dummy variables for each grade.

[b] Regressions for a specific grade stack across the three subjects for that grade and include dummy variables for each subject. The row labeled "All grades" stacks across both subjects and grades and includes dummy variables for each grade/subject combination. The row labeled "All grades with Ttexts" excludes grade/subject combinations that did not received textbooks.

Table A.4: Access to School Provided Textbooks

			Had textbook to use in class?		Had textbook to take home?	
Year	*Subject*	*Grade*	*Textbook schools (%)*	*Comparison schools (%)*	*Textbook schools (%)*	*Comparison schools (%)*
2019	English	6-7	78.9	13.4	67.9	4.9
		8	22.6	15.0	15.9	4.9
	Math	6-7	46.6	11.6	37.2	3.7
		8	12.7	17.2	8.4	5.3
	Science	6-7	2.7	3.3	1.1	1.7
		8	59.8	6.0	54.1	3.0
	All (if books received)		62.4	11.5	52.8	4.1
	All (if no books received)		8.6	8.7	5.4	3.1
2019	English	6-7	74.8	33.7	65.7	13.9
		8	36.7	27.2	28.3	
	Math	6-7	72.8	29.6	64.1	12.9
		8	33.9	22.8	27.2	
	Science	6-7	16.7	9.7	11.4	4.7
		8	59.7	12.3	54.2	
	All (if books received)		72.0	29.0	63.5	12.3
	All (if no books received)		23.4	15.6	17.4	7.9

A5: Availability of Textbooks
(book per pupil ratios)

			School books		*Private books*		*School+Private books*	
Year	*Subject*	*Grades*	*Text. schs.*	*Comp. schs*	*Text. schs.*	*Comp. schs*	*Text. schs.*	*Comp. schs*
	English	3-7	0.67	0.03	0.09	0.15	0.76	0.19
		8	0.11	0.11	0.36	0.28	0.47	0.39
	Math	3, 5, 7	0.61	0.05	0.09	0.15	0.70	0.20
		4, 6	0.04	0.03	0.09	0.15	0.14	0.18
		8	0.06	0.10	0.36	0.30	0.42	0.40
9								
9	Science	3-7	0.01	0.02	0.01	0.02	0.02	0.04
6		8	0.68	0.05	0.13	0.09	0.81	0.14
	All (if books received)	3-8	0.65	0.04	0.09	0.15	0.74	0.19
	All (if no books received)	3-8	0.02	0.03	0.05	0.07	0.08	0.09
	English	3-7	0.54	0.04	0.09	0.18	0.62	0.22
		8	0.14	0.08	0.35	0.36	0.50	0.44
	Math	3-7	0.55	0.04	0.09	0.16	0.64	0.21
1		8	0.13	0.08	0.36	0.35	0.50	0.43
9								
9	Science	3-7	0.02	0.02	0.02	0.02	0.05	0.04
7		8	0.56	0.04	0.20	0.13	0.76	0.17
	All (if books received)	3-8	0.55	0.04	0.09	0.17	0.64	0.21
	All (if no books received)	3-8	0.04	0.03	0.08	0.08	0.12	0.12
1								
9	English	3-7	0.55	0.12	0.08	0.15	0.64	0.27
9		8	0.29	0.21	0.43	0.45	0.72	0.66
8								
	Math	3-7	0.47	0.10	0.08	0.12	0.56	0.22
		8	0.21	0.15	0.40	0.40	0.61	0.56
	Science	3-7	0.08	0.08	0.03	0.03	0.11	0.11

		8	0.62	0.13	0.22	0.18	0.84	0.31
	All (if books received)	3-8	0.52	0.11	0.08	0.14	0.61	0.25
	All (if no books received)	3-8	0.11	0.09	0.09	0.09	0.20	0.19

Note: Comparison schools varied by year. In 1996 it was the 75-school comparison group, while in 1997 and 1998 it was the 50-school and 25-school comparison groups, respectively.

Table A6: Textbook Use and Classroom Activities - 1997

	N	*Mean treat. (%)*	*Mean comp. (%)*	*p-value of difference*
CLASSROOM OBSERVATION:				
Textbook used in class?	461	68.4	52.1	0.332
Does the teacher have a plan book?	359	13.2	21.1	0.106
Teacher's energy level [1: least enthusiastic – 5: most enthusiastic]	356	3.26	3.18	0.060
Pupil participation [1: least enthusiastic – 5: most enthusiastic]	365	3.25	3.10	0.004
Teacher organized [1: not organized – 3: well organized]	364	1.55	1.56	0.374
Teacher activities as % of total time:				
Absent [out of classroom or out of school]	582	31.9	43.7	0.001
Teaching [explaining lesson, asking and answering questions, reading, or writing on blackboard]	582	38.9	34.3	0.080
Assigning or evaluating an exercise, a test or homework	582	15.6	11.9	0.004
Monitoring, disciplining, doing administrative work, distributing exercise books	582	11.7	8.1	0.001
% time teacher used the blackboard as visual	582	21.5	22.0	0.780
% time teacher used the textbook	582	25.6	21.4	0.072
Pupils activities as % of total time:				
Not doing anything [sitting quietly, unfocused, talking to each other, doing what they want]	582	26.3	36.2	0.001
Involved in the lesson [listening, asking/answering questions, reciting, talking about lesson]	582	34.7	29.9	0.066
Using the textbook [copying or reading from the textbook]	582	6.7	5.5	0.136
Doing or evaluating an exercise, taking a test	582	22.2	20.2	0.212
Using the blackboard [copying or reading from the blackboard, writing notes on the blackboard]	582	5.7	5.1	0.430
PUPIL INTERVIEWS:				
Teacher's presence in class [1: never – 4: always]	2889	2.84	2.78	0.424

Used the textbook [read or did exercises from it]	3060	60.8	55.7	0.114
Used the blackboard [copied notes from it, did exercises from it]	3060	30.5	37.3	0.134

*: 10% sig. **: 5% sig. N: number of observations. P-values based on regression analysis. In addition to a treatment dummy variable, all regressions control for geographical location, grade and subject. School level random effects are used in the RE and RPRO regressions.

Table A7: Textbook Use and Classroom Activities - 1998

	N	*Mean treat. (%)*	*Mean comp. (%)*	*P-value of difference*
CLASSROOM OBSERVATION:				
Textbook used in class?	470	43.2	36.1	0.064
Does the teacher have a plan book?	318	45.2	37.5	0.459
Teacher's energy level [1: least enthusiastic – 5: most enthusiastic]	331	3.3	3.2	0.284
Pupil participation [1: least enthusiastic – 5: most enthusiastic]	331	3.1	3.0	0.612
Teacher organized [1: not organized – 3: well organized]	332	1.71	1.72	-
Teacher activities as % of total time:				
Absent [out of classroom or out of school]	463	30.7	50.7	0.000
Teaching [explaining lesson, asking and answering questions, reading, or writing on blackboard]	463	39.6	29.0	0.002
Assigning or evaluating an exercise, a test or homework	463	16.3	11.4	0.000
Monitoring, disciplining, doing administrative work, distributing exercise books	463	11.2	7.5	0.030
% time teacher used the blackboard as visual	463	22.1	18.8	0.318
% time teacher used the textbook	463	21.9	15.1	0.001
Pupils activities as % of total time:				
Not doing anything [sitting quietly, unfocused, talking to each other, doing what they want]	463	25.5	37.6	0.000
Involved in the lesson [listening, asking/answering questions, reciting, talking about lesson]	463	35.4	25.1	0.001
Using the textbook [copying or reading from the textbook]	463	7.0	6.8	0.638
Doing or evaluating an exercise, taking a test	463	22.4	23.1	0.674
Using the blackboard [copying or reading from the blackboard, writing notes on the blackboard]	463	5.9	4.7	0.107
PUPIL INTERVIEWS:				
Teacher's presence in class [1: never – 4: always]	1935	2.92	2.85	0.000

Used the textbook [read or did exercises from it]	1986	66.0	54.0	0.034
Used the blackboard [copied notes from it, did exercises from it]	1986	43.0	70.1	0.000

*: 10% sig. **: 5% sig. N: number of observations. P-values based on regression analysis. In addition to a treatment dummy variable, all regressions control for geographical location, grade and subject. School level random effects are used in the RE and RPRO regressions.

References

Bryk, Anthony, and Stephen Raudenbusch (1992), *Hierarchical Linear Models.* Sage Publications: Newbury Park, CA.

Burtless, Gary. 1996. *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success.* Brookings Institution. Washington, DC.

Card, David, and Alan B. Krueger. 1992. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy* 100(February):1-40

Case and Deaton. 1999. QJE.

Deolalikar, Anil B. 1999. "Primary and Secondary Education in Kenya: A Sector Review."Revised February 1999.

Finn, Chester E., Diane Ravitch and Robert T. Fancher. (1984) *Against Mediocrity : the Humanities in America's High Schools.* New York : Holmes & Meier

Finn, Jeremy and Charles Achilles (1990), "Answers and Questions about Class-Size: a State-Wide Experiment," *American Educational Research Journal* (Fall), 557-577.

Fuller, Bruce (1986), *Raising School Quality in Developing Countries: What Investments Boost Learning?* The World Bank, Washington D.C.

Fuller, Bruce and Prema Clarke (1994), "Raising School Effects While Ignoring Culture? Local Conditions and the Influence of Classrooms, Tools, Rules and Pedagogy", *Review of Educational Research*, 64(1), 119-157.

Glewwe, Paul, Lakshmi Iyer and Michael Kremer. 2002. "School Funding and Student Performance: Evidence from Kenya." Draft paper.

Goldstein, Harvey (1987), *Multilevel Models in Educational and Social Research*. Oxford University Press.

Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24(3):1141-1177.

Hanushek, Eric A. (1995), "Interpreting Recent Research on Schooling in Developing Countries," World Bank Research Observer, 10 (August), 227-246.

Heckman, James J. Anne Layne-Farrar, and Petra Todd. 1994. "Does Measured School Quality Really Matter? Understanding the Empirical and Economic Foundation of the Evidence." University of Chicago, Department of Economics, Chicago. Processed.

Heckman, James J. and Jeffrey A. Smith (1995). "Assessing the Case for Social Experiments," *Journal of Economic Perspectives*, 9(2), pp. 85-110.

Heckman, James J., Robert J. Lalonde and Jeffrey A. Smith (1999). "The Economics and Econometrics of Active Labor Market Programs", in O. Ashenfelter and D. Card, eds. *Handbook of Labor Economics Volume 3A*. North Holland. Amsterdam.

Hedges, L. V., Richard Laine, and Rob Greenwald. 1994. "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23(3):5-14.

Heyneman, Stephen P., Joseph P. Farrell and Manuel A. Sepulveda-Stuardo. 1978. *Textbooks and Achievement: What We Know*. World Bank Staff Working Paper No. 298.

Heyneman, Stephen P., Dean T. Jamison and Xenia Montenegro. 1984. "Textbooks in the Philippines: Evaluation of the Pedagogical Impact of Nationwide Investment." *Educational Evaluation and Policy Analysis* 6(2):139-150.

Jamison, Dean, Barbara Searle, Klaus Galda and Stephen Heyneman. 1981. "Improving Elementary Mathematics Education in Nicaragua: An Experimental Study of the Impact of Textbooks and Radio on Achievement". Journal of Educational Psychology 73(4): 556-67.

Kagitcibasi, Cigdem, Diane Sunar, and Sevda Bekman (1993), "Long-Term Effects of Early Intervention," Department of Education, Bogadzdi University, Istanbul, Turkey.

Kremer, Michael, Sylvie Moulin, Robert Namunyu, and David Myatt. 1997. "The Quantity-Quality Tradeoff in Education: Evidence from a Prospective Evaluation in Kenya", unpublished.

Kremer, Michael, Sylvie Moulin, Robert Namunyu. 2003. “Decentralization: A Cautionary Tale”, unpublished.

Krueger, Alan. (1999) “Experimental Estimates of Education Production Functions,” *The Quarterly Journal of Economics* 114(2):497-532

Levin, Henry, and Marlaine Lockheed. 1993. *Effective Schools in Developing Countries.* Falmer Press. Washington, DC.

Lockheed, Marlaine E., and Eric Hanushek. 1988. "Improving Educational Efficiency in Developing Countries: What Do We Know?" *Compare* 18(1):21-38.

Lockheed, Marlaine, and Adriaan Verspoor (1991), *Improving Primary Education in Developing Countries*, NY: Oxford University Press.

Newman, John, Laura Rawlings, and Paul Gertler (1994),"Using Randomized Control Designs in Evaluating Social Sector Programs in Developing Countries," *The World Bank Research Observer*, (July), 181-202.

Olsen, Randall, J. and George Farkas (1990), "The Effect of Economic Opportunity and Family Background on Adolescent Cohabitation and Childbearing among Low-Income Blacks," *Journal of Labor Economics* 8(3).

Ravitch, Diane (2000*). Left Back : A Century of Failed School Reforms* . New York : Simon & Schuster

Ravitch, Diane (1983*). The Troubled Crusade : American Education, 1945-1980*. New York : Basic Books

Selden, Thomas and Michael Wasylenko (1995), “Measuring the Distributional Effects of Public Education in Peru.” In Dominique Van De Walle, and Kimberly Nead eds., *Public Spending and The Poor.* Baltimore: The Johns Hopkins University Press for the World Bank.

Thobani, Matteen, (1983) “Charging User Fees for Social Services: the Case of Education in Malawi,” World Bank Staff Working Paper no.527, (Washington DC).

UNDP. 1990. *Human Development Report.* Oxford University Press. New York.

UNESCO. 1999. *Statistical Yearbook*. UNESCO: Paris, France and Bernan: Lanham, Maryland.
World Bank. 1989. "Sub-Saharan Africa: From Crisis to Sustainable Growth". The World Bank. Washington, DC.

World Bank. 1990. *World Development Report*. Oxford University Press. New York.