Incentives to Learn

Michael Kremer[*]

Edward Miguel[**]

Rebecca Thornton[***]

October 2003

Abstract: The impact of a merit scholarship program for adolescent girls in Kenya is examined in the context of a randomized evaluation. Girls in program schools were informed that if they scored well on a later academic exam their school fees would be paid and they would receive a large cash grant for the next two years. Girls eligible for the scholarship showed large gains in academic exam scores (average gain 0.2-0.3 standard deviations), and these gains persisted into the year following the competition. There is also evidence of positive externalities: girls with low baseline test scores (with less chance at the award) and boys (who were ineligible) showed sizeable test gains. Both student and teacher school absenteeism fell in the scholarship schools, but there is no evidence of changes in students' self-perceptions or attitudes toward school.

[*] Department of Economics, Harvard University, The Brookings Institution, and NBER. Department of Economics, Littauer 207, Harvard University, Cambridge, MA 02138, USA; 1 (617) 495-9145, mkremer@fas.harvard.edu.
[**] Department of Economics, University of California, Berkeley and NBER. Department of Economics, 549 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, USA; 1 (510) 642-7162, emiguel@econ.berkeley.edu.
[***] Department of Economics, Harvard University, Littauer 207, Cambridge, MA 02138, USA; rlthornt@fas.harvard.edu.

1. Introduction

While much research on education focuses on the effects of material inputs, teachers, or school organization, the most important input in education may be children's time and effort. Yet the impact of performance incentives for students has been relatively unexplored in both economics and education.[1] Even in the presence of large future returns to education in the labor market, student effort may be sub-optimally low (in a lifetime sense) since children have much higher rates of time discounting than adults (Greene 1986, Nurmi 1991).[2]

This paper examines the impact of the Girls Scholarship Project (GSP), a program introduced in rural Kenya in early 2001, on primary educational outcomes. The project provided scholarships for 13-15 year old girls amounting to approximately US$38 per winner over two years – a large sum in this region where annual per capita income is only US$328 (World Bank, 2002). The fact that schools eligible for scholarships were randomly selected allows us to attribute the differences in educational outcomes – primarily academic test scores and school attendance – between the program and comparison groups to the scholarship program. We also use student questionnaire data on study patterns, self-esteem, and attitudes toward school in an attempt to identify mechanisms through which incentives impact education.

There are three main empirical findings from the first two years of the program. First, girls eligible for the scholarship had significantly higher test scores than comparison girls who did not compete for the scholarship, and this result is nearly identical over two rounds of competition for the award. Second, effects of the program on test scores persisted in the year after the competition, even when girls were no longer competing for scholarships. That is to say, there appear to be medium-run benefits of the program that are not simply the result of cramming for the exam. Third, there is some evidence of educational externalities: there are significant improvements in test scores for girls with low initial scores, who had less chance at the award, and for boys, all of whom were ineligible for the scholarship.

---

[1] This literature is discussed below. An important recent exception is the working paper by Angrist and Lavy (2003).
[2] Both social spillovers to education, or imperfect altruism within the household, could also lead to systematic underinvestment in education.

In terms of mechanisms, scholarship schools had significantly higher rates of both student and teacher school attendance after the program was introduced. Girls in scholarship schools were also more likely to use textbooks at home following the competition for the scholarship (although not more likely to own additional educational inputs such as books or exercise books). There was no significant change in self-expressed attitudes toward schooling or students' own academic ability in the scholarship schools, however, suggesting that social psychological explanations are unlikely to be driving the estimated test score effects.

The findings contribute to at least three literatures. First, they speak to ongoing policy debates in education circles in both wealthy and less developed countries regarding the desirability of academic merit awards. The findings also contribute to the literature in education and psychology on the impact of extrinsic versus intrinsic motivation, as we discuss more below; our finding that the introduction of the scholarship did not reduce the performance of either the eligible students (girls) or ineligible students (boys) in the program schools suggests that extrinsic motivation was not harmful for school performance in this context. Finally, the results provide evidence for educational externalities, and possibly classroom peer effects.

The remainder of the paper proceeds as follows: Section 2 provides background information on schooling in Kenya and the Girl's Scholarship Program that we evaluate. Section 3 discusses the literature on extrinsic rewards in general and merit scholarships in particular. Section 4 discusses the data set, sample attrition, and presents the estimation strategy. Section 5 discusses the main empirical results, and the final section concludes.

2. Merit Scholarships and Academic Performance
2.1 Evidence on Merit-Based Scholarships

Historically, many scholarships in the United States were merit-based, but over time there has been a move toward need-based scholarships. Recently, however, there has been a resurgence in popularity of merit scholarships. While more than three-quarters of all state funded scholarships in the United States are

based on financial need, non-need based funds have grown almost fifty percent in the past five years (The College Board 2002).

A number of studies suggest scholarships increase educational participation (for instance, Dynarski 2003), but fewer studies examine the incentive effects of merit scholarships. Binder et al. (2002) find that while scholarship eligibility increased students' grades, the number of credit-hours completed decreased, suggesting that students took fewer courses in order to keep their grades high. Cornwell et al. (2002) found that the average SAT score for Georgia's high school seniors rose almost 40 points after the HOPE scholarship program was introduced, although this increase may not have been a direct result of the scholarship program itself.

In work most closely related to the current study, Angrist and Lavy (2003) find that cash rewards for test performance raised high school graduation rates in Israel. They examine a pilot study involving 500 high school students, pseudo-randomly selected, and a scholarship program that randomly selected 20 program and 20 comparison schools. Students in program schools were able to earn cash based on their performance in matriculation exams. Students offered the merit reward had approximately 8.5 percentage points higher high-school exam completion rates than comparison schools, with the largest positive effects of the program among the upper quartile of students.

The program evaluated in Israel differs from the one evaluated in this paper in several important ways. First, because of a number of political and logistical concerns with the scholarship program in Israel, the program, which was meant to last for three years, was discontinued after the first year, making it impossible to estimate longer-term impacts, and effects once the incentive has been removed. Second, Angrist and Lavy examine just 20 treatment schools and 20 comparison schools, while our study involves more than three times as many schools. Third, the sample of students randomly assigned to treatment in the Israel study was not large enough to insure that average characteristics in control and treatment schools were similar, and Angrist and Lavy therefore use non-experimental matching techniques to pair treatment and control schools based on previous exam completion rates.

The current study, on the other hand, has a balanced program and comparison group prior to the announcement of the scholarship (as discussed below). In addition to test score results, we also collected original survey information on student school attendance, teacher school attendance data, and a range of student attitude and behaviors (including self-esteem, and study habits at home), and this allows us to explore possible mechanisms through which a merit scholarship affects students. Documenting and understanding these channels is particularly important because such programs may generate incentives to manipulate test scores by teaching to the test, for instance, or cheating (for a discussion, see Glewwe, Ilias, and Kremer 2003).

In a related study, Ashworth et. al (2001) study the impacts of financial support in the form of weekly allowances given to 16-19 year old students from low-income households in England, with some additional bonuses based on retention and academic achievement. In a pilot study evaluation of Education Maintenance Allowances (EMA), findings indicate that the program raised school participation among eligible youth by 5.9 percentage points and among both ineligible and eligible youth by 3.7 percentage points, suggesting program externalities. Although the study utilized longitudinal data from two cohorts, they employ propensity score matching based on observable demographic characteristics to estimate impacts, since program areas were not randomly selected; rather, the program was targeted at urban areas known to have high levels of deprivation. In a similar study, Croxford et. al. (2002) evaluate Education Maintenance Allowances in Scotland and find increases in post-compulsory education by 7 percentage points. However, it remains unclear how much of these impacts are due to rewarding students for school participation versus for their academic achievement.

Anecdotal evidence suggests that merit based scholarships are effective in providing positive reinforcement to students and are successful in raising high school graduation rates, test scores, and attendance, and this view has led a number of philanthropic organizations to provide merit scholarships for both high school and college to students in poor U.S. neighborhoods in the United States.

2.2 A Model of Incentives and Study Effort

A stylized economic framework helps to illustrate the potential impact of merit scholarship. We model the academic test score, $TEST_{ist}$, as a function of the school effort, e_{ist}, of student i in school s in the current period t. This effort may take various forms, including improved school attendance or working harder on homework. The test score may also be a function of teacher effort, e^T_{st}, where this can take the form of improved teacher attendance or effort in preparing lessons; of the average schooling effort of other children in the class, e^C_{st}, a classroom peer effect, since it may be easier to learn in a setting where other children are also exerting effort and are serious about their studies; as well as a function of the child's current ability, A_{ist}, which is a function of past effort exerted by the child herself, by her teachers, and by her other classmates, as well as of the innate ability of the child (A_{is0}).[3] Past effort (for time t-1, t-2, …) of the student herself is denoted by the vector $e_{is,-1}$, with analogous notation for past effort by teachers and classmates:

(1) $$A_{ist} = A(e_{is,-1}, e^T_{is,-1}, e^C_{is,1}, A_{is0}).$$

In practice, ability may be difficult to directly observe, although academic test scores serve as a proxy. The test score is thus formally determined as follows:

(2) $$TEST_{ist} = f(e_{ist}, e^T_{ist}, e^C_{ist}, A_{ist})$$

where the test score is increasing in each argument of the function. It is possible that the effort of children and their teachers, and of children and their classmates are either complements (formally, that $f_{12} > 0$ and $f_{13} > 0$, respectively) or substitutes, and we do not impose either in the model. Similarly, own effort and ability may be either complements or substitutes contemporaneously, and own effort at one point in time may complement or substitute effort at other points in time (working through the ability term). If effort and ability are complements, for instance, then a one-time increase in study effort that boosts A_{ist} – through the scholarship incentive – could lead to persistent future increases in study effort.[4]

---

[3] We need not impose the restriction that A_{ist} is a sufficient statistic for previous educational inputs, although this simplifies the model somewhat.

[4] Note that other possible channels for persistent effects of the program is the payment of school fees through the scholarship program in future years, as well as income effects from the cash transfer to winners.

The Girls Scholarship Project that we study directly affects incentives to exert school effort. Academic effort first leads to general future wage and non-pecuniary benefits, *B*, which is a concave function of the same factors that determine the test scores. Effort also increases the perceived probability, *P*, that an individual will win the scholarship, which has total monetary and non-monetary value $S > 0$. The non-monetary benefits from the GSP include the public recognition winners received in a school assembly. The probability of winning a scholarship is a function of the same factors that affect the test scores, and of the assignment to treatment or comparison groups of school *s*, T_s, where T_s takes on a value of one for scholarship schools. The probability of winning the scholarship is zero for all students in comparison schools, as well as for boys (and girls in grades other than Grade 6) in scholarship schools. Finally, the cost of exerting academic effort is the convex increasing function *C*. Thus the maximization problem facing students takes the following form:

$$(3) \qquad \underset{e_{ist}}{Max}\, B(e_{ist}, e_{st}^T, e_{st}^C, A_{ist}) + S \cdot P(e_{ist}, e_{st}^T, e_{st}^C, A_{ist}, T_s) - C(e_{ist})$$

The introduction of the award leads to greater school effort among those eligible for the award, and among those who will be eligible in future years (since they seek to increase their ability, A_{ist}). The award would also lead to persistent test score gains, since the one-time increase in effort made in order to win the award raises future ability, A_{ist}. A simple extension implies that teachers in scholarship schools should also exert more effort than teachers in comparison schools. If teachers face a similar maximization to that presented above, in which they experience some non-monetary (or even monetary) gain from having more scholarship winners in their class, then they should increase their work effort in order to boost student performance. Part of the non-monetary cost of shirking for teachers could be pressure, or informal social sanctions, on the part of parents in the community.

This framework illuminates how even those individuals in scholarship schools who are ineligible for awards (i.e., boys) or who are eligible but unlikely to win awards (i.e., girls with very low baseline academic ability) might also experience externality benefits from the program, through at least three possible channels. First, they could directly benefit from the higher teacher effort (e^T_{ist}), which is an input

in their own learning. Second, the greater effort of classmates (e^{C}_{ist}) could lead to a better classroom environment and boost their scores, through a classroom peer effect. Third, to the extent that the student's own effort complements teachers' and classmates' effort, even children without any incentive from the program would find it optimal to exert additional school effort themselves, again boosting their scores.

In the empirical work that follows we focus on reduced form estimation of this model, in other words, on the impact of being assigned to be a scholarship school on academic test score outcomes. Yet we also estimate the impact of treatment assignment on a range of possible channels linking behavior to test scores, in particular, various measures of student and teacher effort, as well as a range of other factors – such as child attitudes toward school and confidence about their own performance – that are not included in the above framework, in an attempt to disentangle the causes of the reduced form effects.

2.3 Evidence on Extrinsic Motivation from Psychology

There are differing views on the potential effects of offering extrinsic rewards to individuals. First, in accordance with a standard economic model such as that above, incentives could increase individual work effort. A second theory, from psychology, asserts that extrinsic rewards may interfere with intrinsic motivation and thus could actually lead to immediate negative effects on work effort. A weaker version of this view is that incentives may indeed lead to greater performance over the short-run, but lead to negative effects over the longer-term, and especially when the incentive is removed, by interfering with intrinsic motivation. Our empirical approach allows us to test these claims, and we ultimately find evidence consistent with the first view. One strength of our approach over many related studies is that we estimate impacts in a real-world context rather than in laboratory experiments.

Early experimental research on learning supported the idea laid out above that reward-based incentives lead to increased effort in students and became a foundation for "positive reinforcement" teaching methodologies (Skinner 1961). Initial laboratory research conducted in the 1970's studied changes in behavior in individuals completing tasks before and after receiving extrinsic motivational

reward and found evidence that there were some situations where external rewards seemed to have

negative impacts (Deci 1971; Lepper et al. 1973).[5]

Later work has typically consisted of laboratory studies that compare participants in groups

completing tasks with and without extrinsic rewards and attempts to quantify the effects of extrinsic

motivators on intrinsic motivation; these have yielded mixed conclusions. Cameron and Pierce (1994) and

Cameron, Banko et al. (2001) conducted meta-studies of over 100 experiments and found that the

negative effects of external rewards were limited and could be easily overcome in several applied settings,

such as for high-interest tasks or when rewards were specifically linked with behavior or measurable

achievements. On the other hand, Deci, Koestner et al. (1999) find strong negative effects of rewards on

task interest and satisfaction in a similar meta-study, so the literature appears ultimately inconclusive.

3. Girls Scholarship Program in Rural Kenya

3.1 Schooling in Kenya

The Girls Scholarship Program (GSP) was carried out by a Dutch non-governmental organization (NGO),

ICS Africa, in Busia district and Teso district, two rural districts in western Kenya. Busia district is

mainly populated by a Bantu ethnic group of agricultural traditions, the Luhya, while Teso district is

populated primarily by a Nilotic ethnic group of pastoralist traditions, the Tesos. These ethnic groups

differ in language, history, and certain present-day customs, although they do not differ greatly along

observed socio-economic characteristics, such as average years of education or asset ownership.

Schooling in Kenya consists of eight years of primary school and four years of secondary school.

While most children enroll in primary school, there are large dropout rates in grades 5, 6, and 7, and only

a fraction of students continue on to secondary school. The dropout rate is especially high for girls[6].

Grade 8 students take the official Kenya Certificate of Primary Education (KCPE) exam. Continuation to

---

[5] Benabou and Tirole (1999) build a principal-agent model in which rewards may undermine self-confidence if the principal offering praise or rewards is more knowledgeable about the agent's ability than the agent herself. However, the fact that all female students in Grade 6 were eligible for the award in our case, and that the award did not rely on the NGOs discretion, implies that this effect is unlikely to apply in our case.

secondary school is contingent on performance on the KCPE and thus students take this exam quite seriously. To prepare for these exams, students in grades 4-8 typically take standardized exams (called "mocks") at the end of each school year, although these exams are sometimes canceled, for example, due to teacher strikes or fears of violence prior to national elections. End-of-year exams are standardized throughout each district and test students in English, Math, Swahili, Science, and Geography-History-Christian Religion (GHC).[7] Students must pay a fee to take the exam, and we discuss implications of this fee below.

During the first year of the scholarship program, students had to pay school fees, levied by the local school committee and headmaster. In late 2001, Kenya's president at the time, Daniel Arap Moi, announced a ban on school fees, but the policy was unclear and unenforced, and school committees at most primary schools continued to charge yearly fees that were typically approximately US$6.4 (500 KSh)[8] per family each year. In practice, these fees set a benchmark for bargaining between headmasters and parents, and many parents did not fully pay these school fees. In addition to the per family school fee, there are also fees for particular activities, such as taking mock exams (as noted above).

Mwai Kibaki became the new president of Kenya following elections in December 2002, and a new policy of free primary education announced during his campaign was put into place in early 2003. Despite this change, students' families must still pay the cost of uniforms, books and supplies for their children, where the average school uniform costs approximately US$6.4 dollars (500 KSh). This policy change took place after the period that we study (March 2001 to February 2003).

3.2 Project Description

The Girls Scholarship Program (GSP) was announced in March 2001. Out of a set of 127 primary schools in the two districts, half were randomly chosen to be invited to participate in the program. The randomization was stratified by political divisions (there are eight divisions in Busia and Teso districts),

---

[6] Girls in the baseline sample enrolled in January 2001 had a dropout rate of 9 percent through late 2001.
[7] Students may also elect to take Islamic Religious Education (IRE). Most students in western Kenya, however, are Christian and take GHC.

and along participation in a past NGO assistance program, which provided wall charts to the schools.[9]

Randomization was done using a computer random number generator, and as we discuss below (Section 4), this procedure was successful at creating treatment and comparison groups largely similar along baseline characteristics.

The scholarship program provided winning girls in Grade 6 with an award for the next two academic years, Grades 7 and 8. The award consisted of the following each year: (1) a grant of US$6.4 (500 Kenyan shillings, KSh) intended for the winner's school fees and paid directly to her school; (2) a grant of US$12.8 (1000 KSh) paid to the girl's family and intended for school supplies and uniforms[10]; and (3) public recognition and acknowledgement at a school awards assembly conducted by the NGO. There may also be benefits for winners' siblings, due to both the income transfer to their family and because primary school fees in western Kenya are levied per household, rather than per student, so the cost of school declined for them as well.[11]

Given that many parents would not otherwise have fully paid school fees, primary schools with winners benefited from the grant money that paid for winners' school fees.[12] Some of these funds may have benefited headmasters and teachers, for example, if they were used to improve the staff room. Moreover, Grade 6 teachers received much of the credit locally for the success of girls in their class.

Under the program two cohorts of Grade 6 girls competed for scholarships based on test scores. Girls registered for Grade 6 in January 2001 in program schools were the first eligible cohort and those registered for grade 5 in January 2001 made up the second cohort. The NGO restricted eligibility to those girls who were already enrolled in the school in January 2001, before the program was even announced.

---

[8] One US dollar was worth approximately 78.5 Kenyan shillings in January 2002 (Central Bank of Kenya, 2002).

[9] Schools participating in the wall chart program in 1997 had not been recipients of previous textbook and grant programs; these schools are quite typical of local schools in terms of pre-intervention characteristics – see Glewwe et al. (2003) for details on the ICS wall chart program.

[10] While specific spending of the grant money was not monitored, parents and head teacher were told publicly that the money was to be spent on educational supplies. Informal interviews with head teachers and several winning girls indicated that the money, at least in part, did in fact go towards purchasing items such as books, new uniforms, math sets, and watches for school.

[11] We will examine effects on the school participation of younger siblings in future research.

[12] Although mandatory school fees were abolished in 2003, as described above, the scholarship program continued to pay grant money directly to schools with scholarship winners.

Thus there was no incentive for students to transfer into program schools. In fact, only approximately five percent of sample students were reported to have transferred schools between January 2001 and January 2002, and incoming transfer rates are nearly identical in the scholarship and comparison schools.

In November 2000, cohort 1 students took their end-of-year Grade 5 exams, and these are used as baseline test scores in the evaluation.[13] In March 2001, the NGO held meetings with the head teachers of schools chosen for the scholarship program to inform them of program rules and to give each school the choice to participate. Head teachers were asked to relay the information about the program to their students by means of a school assembly. However, there was apparently considerable variation in the extent to which head teachers disseminated this information about the scholarship, so the NGO held additional community meetings in September and October 2001 After these meetings, NGO enumerators began collecting student attendance information during unannounced school visits.

In June 2001, a tragic accident occurred in which lightning struck a school in Teso district, killing seven students and injuring 27 others. Because ICS had been involved with other assistance programs in that school, and due to certain strange coincidences (the names of certain lightning victims were the same as the names of ICS staff members who had recently visited the schools) the deaths were associated with ICS in the eyes of some community members, and several schools requested to withdraw from participation in the Girls Scholarship. Out of the original 58 schools selected for the sample in Teso district, five pulled out of the program. In addition one school in Busia that was close to the Teso border also pulled out. Half of the schools that pulled out of the program were treatment schools. We discuss the implications for econometric inference below.

Students took Grade 6 mock exams in November 2001, and each district, Busia and Teso, gave their own exam. Scholarship winners were chosen based on their scores in subject test scores in English, Swahili, Math, Science, and Government-History-Christian Religion. The NGO then awarded scholarships to the highest scoring 15 percent of Grade 6 girls in the program schools in each district (this

---

[13] A detailed project timeline is presented in Appendix Table A.

translated into 110 girls in Busia district and the 90 girls in Teso district). Schools varied considerably in the number of winners, with almost 60 percent of program schools having at least one scholarship winner in 2001. Among schools with at least one winner, there were an average of five winners per school, while the best performing school had a total of 24 winning girls.

In January 2002, the NGO held school assemblies to announce the 2001 winners. Students, teachers, parents and district officers were invited to the assemblies to recognize each winner publicly. The winner was awarded an ICS certificate, parents received the US$12.8 (1000 Ksh) cash grant, and the school received US$6.4 (500 Ksh) to cover school fees. The community was reminded that the program would continue for another year, and of the program rules. Both the mother and father of winning girls were instructed that the money was to be used to purchase school materials, such as school uniforms, textbooks, exercise books, pens, and pencils. For the rest of the 2002 academic year, the NGO returned regularly to both scholarship and comparison schools to conduct unannounced attendance checks and to administer the 2002 Student Questionnaires to students in Grades 5-7, which collected information on schooling effort and attitudes (as described below).

Exams were again held in late 2002 in Busia district, but because of threatened disruption due to the Kenya national elections held during this time, the exams in Teso district were canceled, so the NGO administered standardized academic exams itself in February 2003 in both Teso and Busia districts. Thus the second cohort of scholarship winners were chosen in Busia district based on the official 2002 exams, while winners from Teso district were determined from the NGO exams. In this second round of the scholarship, forty-four of the program schools had at least one winner, an increase over the first year.

One concern with the test score data is the possibility of cheating in scholarship schools, but this is unlikely for a number of reasons. First, Kenyan district education offices have a well-established system of exam supervision in place, with exam proctors (called "invigilators") from outside the school monitoring all exams, and punishments for schools where widespread cheating is discovered. Invigilators document and punish all instances of cheating, and report these statistics back to the district office. District records indicate there were no documented instances of cheating during 2001 or 2002 exams.

Another finding argues against the cheating explanation: as we report below, test score gains among cohort 1 students in scholarship schools persisted into the year after the competition, and there were also gains among program school boys – groups for which there would be no motivation to cheat.

4. Data and Measurement

4.1 The GSP Dataset

The test score data used in the econometric analysis were collected from the District Education Officers (DEO) in Busia and Teso districts. Test scores were normalized by academic subject in each district, such that the normalized scores in comparison schools are distributed with mean zero and standard deviation one. We use data for cohort 1 and cohort 2 students in the main analysis, for students with age, school attendance, and gender information, in schools that did not pull out of the program and for which we have baseline test scores from 2000 and school ethnic composition (Table 1). We call this the restricted sample. Note that average test scores are generally somewhat higher in this restricted sample than in the baseline sample, since the students dropped from the unrestricted sample are typically somewhat below average in terms of academic achievement, as discussed more below. We utilize school average test scores in 2000, the year before the introduction of the program, as a baseline control in most specifications (Panels B and C).

Attendance data are based on unannounced checks made in September or October 2001, and then throughout the 2002 academic year, by NGO enumerators, who recorded students from the initial school roster present on the day of the check as "present". Note that attendance rates among sample students are roughly 80 percent. These data are an improvement over data from official school registers, which are often thought to be unreliable in less developed countries.

Household demographic and socioeconomic characteristics are similar across scholarship and comparison schools (Table 2): there are no significant differences in terms of parent education, number of siblings, or the ownership of latrines, an iron roof, or mosquito nets (using data from the 2002 student surveys), indicating that the randomization was largely successful in creating *ex ante* comparable

groups.[14] Further evidence is provided by comparing the 2000 (baseline) test score distributions, which are nearly identical graphically (Figure 1). More formally, we cannot reject the hypothesis that average test scores are identical across scholarship and comparison schools among program non-attritors (Table 4, column (vi) – discussed more below).

4.2 Sample Attrition

Sample attrition was extensive in Teso district schools, but not in Busia district. Of the 127 schools invited to participate, six refused, leaving 121 schools. Six additional schools (four in Teso district) with incomplete 2000 exam scores, or 2001 or 2002 exams, or missing demographic age and ethnicity data were also dropped, leaving 115 schools and 6983 students in both cohorts of the *restricted sample*, of which students in the scholarship schools account for exactly 50 percent. Of the twelve schools that are not included in the restricted sample, nine were in Teso district, accounting for 16 percent of the original Teso district sample (while only 4 percent of the original Busia district schools were dropped).

There is a large and statistically significant difference in attrition across scholarship and comparison schools in Teso district schools, but there is much less of problem in Busia. In Teso, among cohort 1 students, 63 percent of scholarship school students took the 2001 exam, while the rate for comparison school students is 77 percent (Table 3, Panel B1). There are several potential explanations for the large and asymmetric attrition between scholarship and comparison schools in Teso district. According to ICS enumerators and local school head teachers, the most important factor is the long-standing misunderstanding between the NGO and some communities in Teso district, in part driven by a general mistrust of outsiders in that district. Although ICS made attempts to abate misconceptions through parent and community meetings, many individuals had long been suspicious of the intentions of outside NGO programs in their communities. The tragic lightning incident mentioned above was literally the spark that set off additional hostility to the NGO, leading several schools to pull out of the program. As

---

[14] This comparison relies on the assumption that the household characteristics (i.e., parent education, fertility, and asset ownership) were not directly affected by the scholarship program by mid-2002, which seems reasonable.

further evidence of the distrust, one girl in Teso district who won the ICS scholarship in 2001 refused the scholarship money.

Thus while only four percent of students in Busia district were in schools that pulled out of the program, fully 12 percent of students in Teso were in schools that left the program (Table 3, Panels C1 and C2). In Busia, there is much less evidence of differential attrition across program and comparison schools. Among cohort 1 students, 82 percent of baseline students in scholarship schools and 77 percent of baseline students in comparison schools in Busia district took the 2001 exam. Thus there is a small, positive but insignificant point estimate, 0.04, on the difference between the proportion taking the 2001 exam between scholarship and comparison schools (Table 3, Panel B1). Among cohort 2 students the difference is even smaller: 52 percent of scholarship school students and 51 percent of comparison students in Busia district took the 2002 exam (Panel B2). There is more overall attrition by 2002 because students drop out of school over time.

Differential attrition between program and comparison schools in Teso district is smaller among cohort 2 students in 2002. To understand why rates of exam taking were higher, and why there is minimal differential attrition, recall that the 2002 district exams for Teso were canceled in the run-up to Kenyan national elections. The NGO instead administered their own exam, modeled on standard government exams, in Teso in 2003, and students did not need to pay a fee to take this exam, unlike government tests, and this is likely to account, at least in part, for the low levels of attrition in Teso district schools for cohort 2.

Among cohort 1 students, the restricted sample includes 73 percent of scholarship school students and 76 percent of comparison students in Busia district (Table 3, Panel D1). In Teso, however, only 54 percent of program students and 58 percent of comparison students remain in the unrestricted sample (Panel D1), thus attrition rates are much higher in Teso district than in Busia, and less balanced in scholarship relative to comparison schools.

There is some evidence that the scholarship program led academically weaker students in program schools who ordinarily would not have paid to take mock exams to do so, potentially biasing

estimated program impacts downward. Theoretically, the introduction of a scholarship could have induced poorer, but high-achieving students to take the exam, leading to an upward bias in the estimated effect of the program, but we do not find any evidence for this. Empirically, examining individual test scores from cohort 1 in 2000 reveals that students who did not take the 2001 exam (i.e., attritors) tended to be somewhat lower achieving students in both Busia and Teso (Table 4, Panel B). Examining the differences in 2000 baseline test scores between attritors and non-attritors shows that Busia program school students who did not take the 2001 exams scored 0.07 standard deviations lower in 2000 than those who did take the 2001 exams, and the difference is 0.56 standard deviations in Busia comparison schools, providing evidence that more low performing students attrit from comparison schools than scholarship schools, and thus suggesting that resulting program impact estimates are likely to be lower bounds on true effects. Below we discuss the construction of bounded estimates that take into account potential bias from attrition.

In reference to the types of students who attrited from the sample due to their schools pulling out of the scholarship program, Teso students in program schools had higher baseline test scores in 2000 than those in comparison schools (Table 4, Panel A). Teso district students whose schools pulled out of the scholarship program, and were in scholarship schools, were typically higher achieving students than those in the comparison schools, scoring a massive 1.48 standard deviations higher in 2000 on average. This is perhaps due to the possibility the high-performing students in program schools were more suspicious of ICS and felt more "vulnerable" to the effects of the program – since they were more likely to win – than those in the comparison schools. In addition to reports from ICS and head teachers in Teso district relating problems within the district, the data thus confirms many of the difficulties in implementing the scholarship program in Teso. Attrition rates in Teso district are high and asymmetric, threatening the validity of estimation in that district, and making standard bounding methods largely uninformative.[15]

---

[15] Bounds of the estimate of the effect of the scholarship program in Teso district following the method in Lee (2002) range from -0.16 to 0.01 standard deviations in the unrestricted sample (for 2001, cohort 1 girls).

For these reasons, we focus our analysis and interpretation on Busia district in what follows, where there was no evidence of differential attrition.

5. Main Empirical Results

5.1 Estimation Strategy

The main estimation equation is:

$$(4) \qquad TEST_{ist} = Z_{ist}' \boldsymbol{b} + (Z_{ist} * T_s)' \boldsymbol{g} + X_{ist}' \boldsymbol{d} + \boldsymbol{m}_s + \boldsymbol{e}_{ist}$$

$TEST_{ist}$ is a student test score outcome for individual i in school s and year t. Z_{ist} is a vector of indicator variables for each cohort and year (cohort 1 in year 1, cohort 1 in year 2, cohort 2 in year 1, cohort 2 in year 2), and T_s is an indicator for being a scholarship school. X_{ij} is a vector that includes other explanatory variables, including the age of the student during the year of scholarship competition, the average school test score in 2000, and the ethnic composition of the school. In some specifications we also examine other school-level socioeconomic controls, and explore the interaction of these characteristics with program participation. The disturbance terms consist of $\boldsymbol{m}_s$, a common school effect capturing common neighborhood or headmaster characteristics, and an idiosyncratic individual term, $\boldsymbol{e}_{ist}$, which may capture unobserved student ability or shocks. Error terms are assumed to be independent across students and schools, but are allowed to be correlated across observations in the same school.

We use to a similar empirical approach to estimate the impact of the program on behavioral channels linking the program to test scores, including student and teacher effort (as measured by school attendance). To estimate the impact of the program across individuals with different baseline scores, we employ a non-parametric regression technique following Fan (1992).

5.2 Academic Test Score Impacts

The scholarship program raised test scores by 0.11 standard deviations (standard error 0.05) overall among boys and girls in 2001 and 2002, in both Busia and Teso districts (Table 5, Panel A), and this effect is statistically significant at 95 percent confidence. Restricting attention to Busia district the overall

17

- -

effect is even larger (0.19, standard error 0.07, Panel B). We next separately estimate effects for girls and boys in different cohorts and years. The program effect is 0.27 standard deviations in the restricted sample of girls in cohort 1 in Busia, and 0.22 for cohort 2 (Table 5, Panel C (1)), and in both cases the effects are significantly different than zero at over 95 percent confidence. To address non-random attrition, we also constructed non-parametric bounds on the treatment effects following Lee (2002). These bounds are tight, ranging from 0.23-0.31 standard deviations in the restricted sample of cohort 1 Busia girls, perhaps not surprising given that there is minimal differential attrition across scholarship and comparison schools (Table 3). The results are also robust to examining the change in test scores between 2000 and 2001 (estimate 0.20, standard error 0.12 – regression not shown).

Other explanatory variables have expected effects. The baseline school average test score in 2000 explains much of the test score in 2001. Being one more year older than other students enrolled in Grade 6 decreases test scores by 0.02 standard deviations (in Kenya, older students within the same grade have usually either repeated a grade or have entered school later than others). Ethnic composition controls generally have weak predictive power. Program impact estimates are similar if these explanatory variables are excluded, although estimates are less precise[16]

However, test score effects in Teso district are near zero and statistically insignificant in both years (Appendix Table B). This is consistent with the hypothesis that winning a scholarship was less desirable in Teso district due to mistrust and fear of the NGO (in the theoretical model, the gains from the scholarship, *S*, are far smaller in Teso district than Busia district schools).

The scholarship program not only significantly raised test scores when it was first introduced in 2001, but also continued to increase scores of cohort 1 students in treatment schools during the second year of the program: the point estimate is 0.24 (standard error 0.08) for the restricted sample, providing additional confidence that the program had lasting effects on learning, rather than simply being due to cheating or cramming for the 2001 exam.

Figures 2, 3, and 4 graph test scores for Busia girls. The vertical line indicates the minimum score needed to win the scholarship. Baseline scores are nearly identical across scholarship and comparison schools. The exam score distribution shifts to the right in program schools for cohort 1 in year 1 (Figure 2), cohort 1 in year 2 (Figure 3), and cohort 2 in year 2 (Figure 4). The largest gains are found near the minimum winning score threshold, on both sides of the score, indicating that the students exerting the most additional effort are those who believe that additional effort will make the greatest difference in their chances to win; those girls at the top of the distribution show positive but smaller gains. This pattern is inconsistent with the hypothesis that teachers cheated by giving all students in their class the correct answers to the test, which would benefit individuals across the score distribution.

From these figures alone it is impossible to determine the magnitude and statistical significance of program effects at different parts of the distribution of initial test scores. Figure 5 presents a non-parametric Fan regression that shows the scholarship program impact for Busia cohort 1 girls based on their individual level test scores in 2000. Girls just below the winning threshold had large test score gains, perhaps due to extra effort exerted in order to win the scholarship. High-achieving girls in 2000 had the smallest increases in 2001 test scores, perhaps since girls with already high scores should exert less effort to win the scholarship, but there are also marked gains at the bottom of the baseline test score distribution, evidence of positive spillover benefits of the program.

Field reports from students in scholarship schools and from head teachers told of how students were encouraged to compete for the scholarship when it was offered. One head teacher reported that the program "awakened our girls and was one step towards making the girls really enjoy school"[17]. One winning girl who was asked about her own performance versus those students who did not win remarked, "they tried to work hard for the scholarship but we defeated them." It is thus plausible that this spirit of competition drove some girls to work harder to win the grant. There are other causes through which test

---

[16] For instance, the program impact for cohort 1 is still exactly 0.27 standard deviations in this case, but the standard error rises to 0.19, while the program impact for cohort 2 rises to 0.31 with standard error 0.18 (significantly different than zero at 90 percent confidence) – regressions not shown.

[17] Source: Authors' field notes, July 15, 2002.

performance for girls may have increased due to the scholarship per se. We analyze these mechanisms in further details below.

Boys in Busia district scholarship schools also have somewhat higher test scores over those in comparison schools, despite not being eligible for the scholarships themselves. Boys' scores in the first year of competition were 0.20 standard deviations (standard error 0.10) in scholarship schools in the restricted sample, despite the fact that the scholarship was only offered to girls (Table 5, Panel C, regression 2). Yet boys and girls in western Kenya share the same teacher and classroom, so any benefits from added teaching effort on the part of teachers could spillover to the boys. It is also possible that some boys were confused whether they were eligible for the scholarship in the first year[18]. In the second year of the program, there are again positive, although statistically insignificant, program impacts for boys. The drop between years 1 and 2 suggests that some boys may have learned that they were not eligible for the scholarship and thus exerted less effort, although we cannot reject that effects for boys are the same in 2001 and 2002. One clear finding, though, is that there were no negative impacts for boys, indicating that the girls' scholarship program did not discourage or demoralize them overall.

The non-parametric Fan regression of the scholarship impact in 2001 for cohort 1 Busia boys based on their baseline level of educational achievement is presented in Figure 6. The patterns for boys are similar to those for girls, with large gains for those boys near and above the scholarship threshold, although the differences at different points of the distribution are generally less pronounced than for girls, consistent with the hypothesis of spillover effects. The large gains among high-achieving boys is possibly due in part to a sense of increased competition with their female classmates.

We next turn to scores by academic subject: the program had large, positive and statistically significant effects on test scores in math and science, and positive but generally smaller gains for English and Swahili, especially for cohort 2 (Table 6). These estimates are broadly similar to the findings of Case

---

[18] We find that boys were less likely to be informed about the scholarship, as well as less likely to understand the eligibility rules. This is discussed in Section 5.3 below.

and Deaton (1999), who found that reductions in student-teacher ratios in South Africa affected math test scores more than comprehensive language test scores.

5.3 Knowledge and Awareness of the Scholarship Program

Awareness of program rules is a prerequisite for the program to have an impact on study effort. The 2002 student survey indicates that girls not only had keen knowledge and awareness of the program, but also may have received parental support in regards to the scholarship. First, we find that students in program schools in both cohorts appear to have decent recognition of the girls scholarship program: 90 percent of students in Busia district claimed to have heard of the scholarship program. General recognition and knowledge levels were typically worse in Teso district schools. A large percent of students' parents (84-92 percent) had also mentioned the program to them.

Girls also had much better knowledge about the specific rules regarding winning the scholarship than boys; for example, girls in cohort 1 were 10 percentage points more likely than boys to know that "only girls were eligible for the scholarship" and eight percentage points more likely to know how many girls in the district would win. Moreover, girls were more likely to report that their parents had mentioned the scholarship program to them at home. This raises the possibility that parents were offering additional encouragement to girls in scholarship schools.

5.4 Student and Teacher School Attendance

There are several possible channels for test score effects, including increased student school attendance, increased study effort at home, increased household purchases of educational materials (e.g., exercise books), increased teacher school attendance, changes in classroom dynamics (including the likelihood that teachers call on students in class), as well as changes in student attitudes regarding education. It is important to understand the channels of program effects, since some mechanisms, such as cheating or increased coaching might raise test scores without improving underlying learning and others, such as shifts. We examine each of these mechanisms in this and the following subsection.

The scholarship program had a significant positive impact on attendance (measured in unannounced enumerator visits) in 2001 and 2002 in Busia district (Table 7, Panel A): pooling cohorts 1 and 2 in the Busia unrestricted sample and measuring the effect of the scholarship on overall attendance yields a coefficient estimate of 0.05 percentage points (0.02 standard error), which is statistically significant at 95 percent confidence, and reduces student absenteeism by approximately 40 percent. These increases in attendance are evidence that students in scholarship schools are exerting extra schooling effort in one important dimension.[19]

The program increased average probability of school attendance by 6 percentage points among girls in cohort 1 in 2001, and by ten percentage points among cohort 2 in 2001 (a pre-program effect), although gains in 2002 are positive but small (Table 7, Panel B, regression 1). The pre-program gains in attendance might be due to anticipation of the future scholarship opportunity, or possibly to peer effects in schools where older girl students in higher grades were competing for the scholarship. Attendance gains were not significantly different between school terms 1, 2 and 3 in 2002 (not shown). Busia district boys also appear to show similar attendance gains in scholarship schools, with large gains in 2001 and essentially no effects in 2002 (Table 7, Panel B, regression 2).

Teachers were five percentage points more likely to be present in program schools during 2002 than those in comparison schools, reducing overall teacher absenteeism by approximately one-third (Table 7, Panel C), and this large effect is robust to alternative specifications (not shown).[20] There are several explanations as to how the program could have increased teacher attendance. Teachers whose students win scholarships are recognized at the school assemblies announcing scholarship winners, they may simply gain pride from their accomplishments, or they may believe that they will receive some other type of recognition in the future. Teachers could also find extra effort more rewarding when their students

---

[19] The estimated student school attendance gain pooling Busia and Teso together is not statistically significant at traditional confidence levels (0.02, standard error 0.02 – not shown), mirroring test score patterns.
[20] Future analysis will estimate the effect on Grade 6 and 7 teachers; the current estimates are for all teachers in the primary school, including those in other grades.

are also putting more into their studies (as a result of the scholarship program).[21] Finally, they may also have been reacting to increased pressure from the head teacher or parents in program schools, given the higher stakes in scholarship program schools.

5.5 Student Attitudes and Behaviors

The 2002 Student Questionnaire collected information on educational inputs, habits, and attitudes that may have led to increases in school performance. Because the questionnaire was administered in 2002, the first cohort of students had already competed for the scholarship when they filled out the questionnaire, while the second cohort of girls had not yet competed.

There are few substantial differences between scholarship and comparison students in either cohort 1 or cohort 2 in terms of their attitudes and behaviors. One notable exception is the finding that there is a significant increase in textbook use among program girls in cohort 1 (Table 8, Panel A): girls in program schools report having used textbooks at home 6 percentage points more than in comparison schools, suggesting that the program led to more intensive studying among the girls (although there is no effect for cohort 2). However, there is no impact on the likelihood that program school students sought out extra school coaching, handed in homework, or were called on in class (Panel A). In terms of educational inputs such as the actual number of textbooks or exercise books available for use at home, there are no significant gains in program schools (Panel B). There is similarly no evidence of any positive or negative impacts on attitudes toward education, for instance, thinking of oneself as a "good student", or preferring school to other activities (Panel C).

We next explore within-school impacts of the program in 2002 – in particular, whether cohort 1 scholarship winners gained in terms of either self-esteem or subsequent test score performance after

---

[21] Schools with winning girls also benefited directly from the scholarship, receiving US$6.4 (500 Ksh) per winner. The scholarship funds represent another channel through which the program could have affected test outcomes in 2002, although given the small amount of these funds (relative to the school budget), and the likelihood that the benefits would be spread across several grades in the school, the impact of these transfers on future test score outcomes is likely to be small.

winning the award – using a regression discontinuity method[22] (Table 9). This method, in effect, compares the 2002 outcomes of girls who barely won the scholarship (their 2001 test score was slightly above the winning threshold) to girls who barely lost out, in order to estimate impacts of winning. In practice, 2001 test score polynomials (linear, quadratic, and cubic terms) are included to control for any smooth underlying relationship between the 2001 test score and 2002 outcomes, and an indicator variable for having a 2001 test score above the threshold then captures the impact of winning the scholarship. By including students in both program and comparison schools, we are able to estimate impacts on both scholarship winners and non-winners, in addition to estimating differences between winners and non-winners in the program schools. The coefficient estimate on the interaction term between the program school indicator variable, and the indicator variable for scoring above the scholarship threshold, captures the impact of winning the scholarship on subsequent outcomes.

Although there is a small positive impact on scholarship winners' view that they are "good students" in both Busia district (Table 9, regression 1), the point estimate is not statistically significant at traditional confidence levels. Note that there is no evidence of a negative impact on academic self-esteem on non-winners in program schools (captured by the coefficient estimate on the program school indicator variable). There is similarly no significant impact of winning the scholarship in 2001 on students' preference for school to other activities in 2002 (regression 2), or their 2002 test score performance (regression 3). In short, although the scholarship program increased school attendance and test score performance, it did not appear to have a pronounced impact on students' academic self-esteem or their enjoyment of school.

6. Conclusion

There is a significant positive effect of a merit-based scholarship for Kenyan girls on test scores in two years of scholarship competition. Reasonably large improvements in test scores are estimated among girls who were initially low-achieving students, and among boys (who were ineligible for the award),

---

[22] See Lee (2001) for further discussion of this methodology.

providing evidence for educational peer effects – rather than the demoralization effect predicted by some psychology theories. Although there are significant effects in Busia district, on the order of 0.2-0.3 standard deviations, we do not find significant effects on test scores in Teso district; this may be due to differential sample attrition that complicates the econometric analysis in Teso district schools, or it may reflect the lower value placed on the program there.

Although the medium-run effects of the program over two years are substantial, it remains an open question how persistent program effects will be on educational, labor market, fertility outcomes in the longer-term. We have collected detailed contact information for individuals in the program and comparison schools, and in future work plan to follow up these individuals as they enter adulthood in order to estimate longer-term program impacts.

References

The College Board (2002). Trends in Student Aid, 2002. Washington, D.C., The College Board.

Angrist, J. and V. Lavy (2003). "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." NBER Working Paper **9389**.

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

Ashworth, K., J. Hardman, et al. (2001). "Education Maintenance Allowance: The First Year, A Qualitative Evaluation". Research Report RR257, Department for Education and Employment.

Benabou, R. and J. Tirole (1999). "Self-Confidence and Social Interactions." Working Paper.

Binder, M., P. T. Ganderton, et al. (2002). "Incentive Effects of New Mexico's Merit-Based State Scholarship Program: Who Responds and How?"

Cameron, J., K. M. Banko, et al. (2001). "Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues." The Behavior Analyst **24**: 1-44.

Cameron, J. and D. Pierce (1994). "Reinforcement, Reward and Intrinsic Motivation: A Meta-Analysis." Review of Educational research **64**(3): 363.

Case, A. and A. Deaton (1999). "School Inputs and Educational Outcomes in South Africa." Quarterly Journal of Economics: 1047-1084.

Cornwell, C., D. Mustard, et al. (2002). "The Enrollment Effects of Merit-Based Financial Aid: Evidence from Georgia's HOPE Scholarship." Journal of Labor Economics.

Croxford, L., C. Howieson, et al. (2002). "Education Maintenance Allowances (EMA) Evaluation of the East Ayshire Pilot." Research Findings No. 6, Enterprise and Lifelong Learnings Report, Glasgow.

Deci, E. L. (1971). "Effects of Externally Mediated Rewards on Intrinsic Motivation." Journal of Personality and Social Psychology **18**: 105-115.

Deci, E. L., R. Koestner, et al. (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." Psychological Bulletin **125**(627-668).

Dickinson, A. M. (1989). "The Detrimental Effects of Extrinsic Reinforcement on "Intrinsic Motivation." The Behavior Analyst **12**: 1-15.

Dynarski, S. (2003). "The Consequences of Merit Aid." NBER Working Paper **9400**.

Fabes, R. A., J. D. Moran, et al. (1981). "The Hidden Costs of Reward and WAIS Subscale Performance." American Journal of Psychology **94**: 387-398.

Fan, J. (1992). "Design-adaptive nonparametric regression." Journal of the American Statistical Association, **87**, 998-1004.

Glewwe, P., S. Moulin, et al. (forthcoming). "Retrospective v. Prospective Analysis of School Inputs: The Case of Flip Charts in Kenya." Journal of Development Economics.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. (2003). "Teacher Incentives", *National Bureau of Economic Research Working Paper #9671.*

Greene, A. (1986). "Future Time Perspective in Adolescence: The present of things future revisited", *Journal of Youth and Adolescence*, 15: 99-113.

Hsieh, Chang-Tai, and Miguel Urquiola. (2003). "When Schools Compete, How do they Compete? An Assessment of Chile's Nationwide School Voucher Program", *NBER Working Paper #10008*.

Kohn, A. (1994). "The Risks of Rewards." Educational Resources Information Center Digest.

Kremer, Michael. (2003). "Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons", *American Economic Review: Papers and Proceedings,* 93 (2), 102-106.

Kruglanski, A., I. Friedman, et al. (1971). "The Effect of Extrinsic Incentives on Some Qualitative Aspects of Task Performance." Journal of Personality and Social Psychology **39**: 608-617.

Lee, D.S. (2001). "The Electoral Advantage to Incumbency and Voters' Valuation of Politicians Experience: A Regression Discontinuity Analysis of Elections to the U.S. House." NBER Working Paper 8441.

Lee, D. S. (2002). "Trimming the Bounds on Treatment Effects with Missing Outcomes." NBER Working Paper.

Lepper, M., D. Greene, et al. (1973). "Undermining children's interest with extrinsic Rewards: A Test of the 'Overidentification Hypothesis." Journal of Personality and Social Psychology **28**: 129-137.

Nurmi, J. (1991). "How do adolescents see their future? A review of the development of future orientation and planning", *Developmental Review*, 11:1-59.

Skinner, B. F. (1961). "Teaching Machines." Scientific America **November**: 91-102.

Figure 1
Distribution of Girl's Test Scores 2000: Busia Restricted Sample Cohort 1
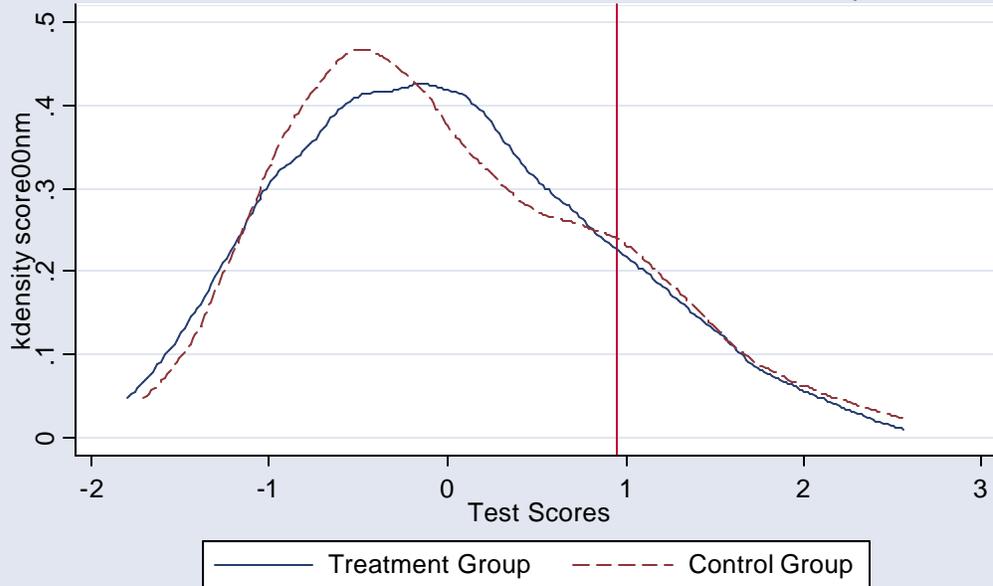
Vertical line represents the minimum winning score in 2001 for girls in Busia. 0.95

Figure 2
Distribution of Girl's Test Scores 2001: Busia Restricted Sample Cohort 1

Vertical line represents the minimum winning score in 2001 for girls in Busia. 0.95

Figure 3

Distribution of Girl's Test Scores 2002: Busia Restricted Sample Cohort 1

Vertical line represents the minimum winning score in 2001 for girls in Busia. 0.95

Figure 4

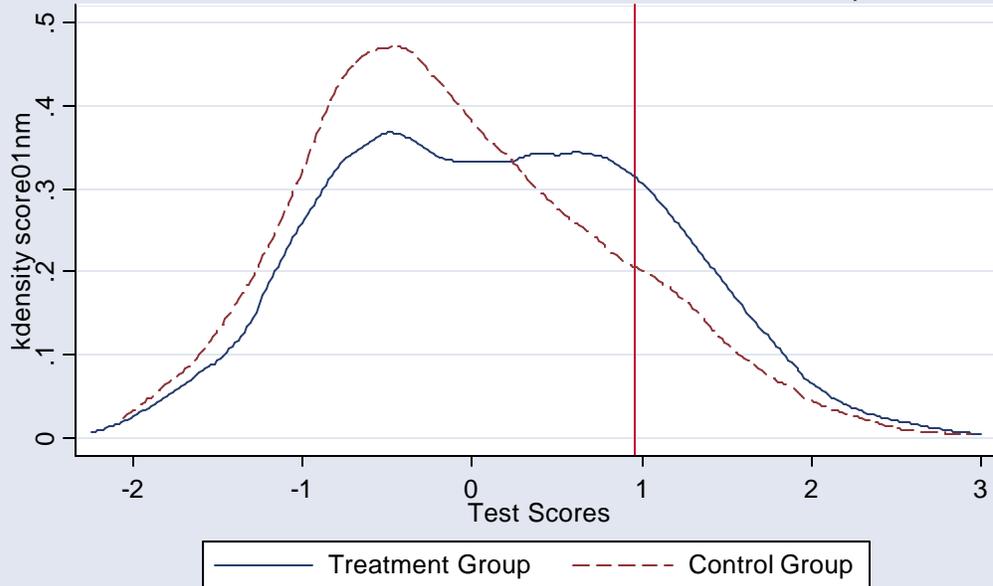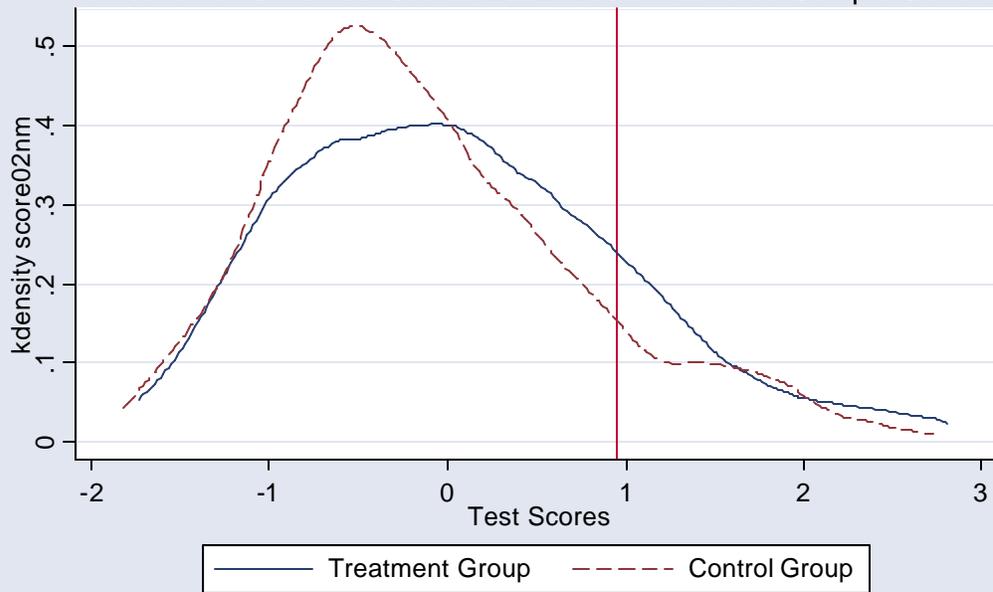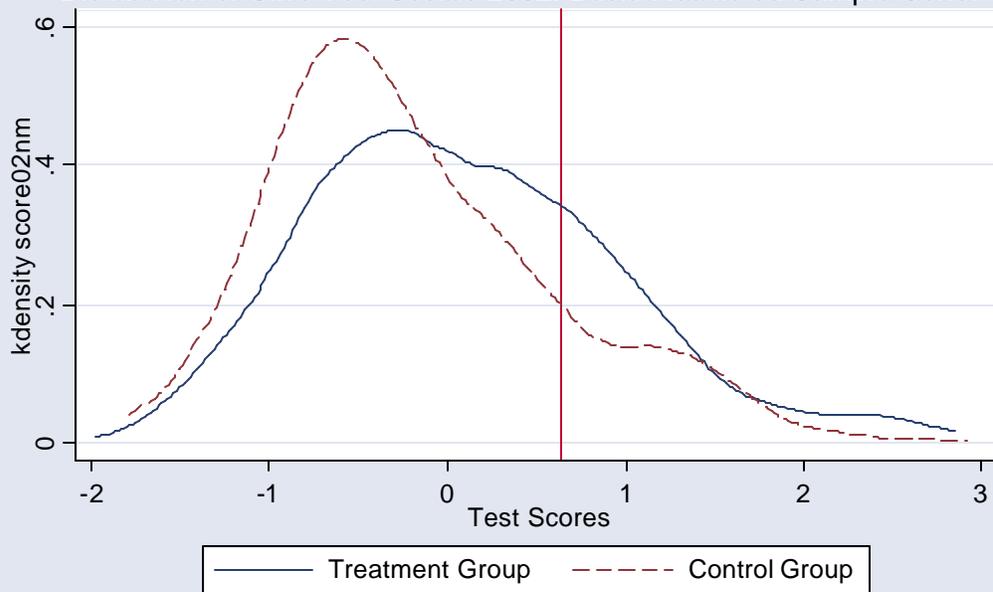### Distribution of Girl's Test Scores 2002: Busia Restricted Sample Cohort 2

Vertical line represents the minimum winning score in 2002 for girls in Busia. 0.64

Figure 5

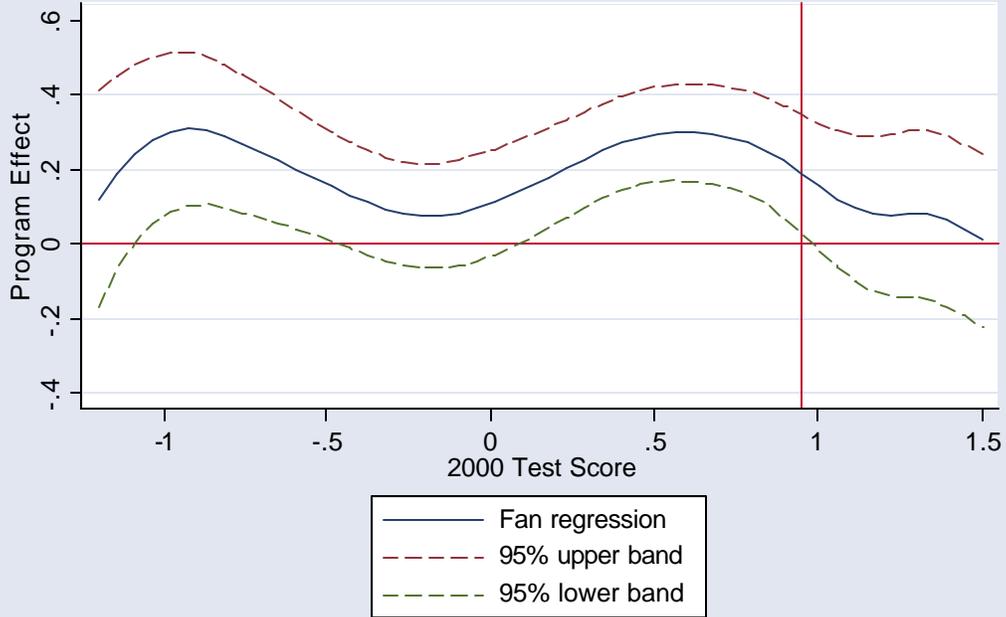Effect of Scholarship on 2001 Test Scores Busia Girls Cohort 1

Fan regression
95% upper band
95% lower band

Figure 6

Effect of Scholarship on 2001 Test Scores Busia Boys Cohort 1

Fan regression
95% upper band
95% lower band
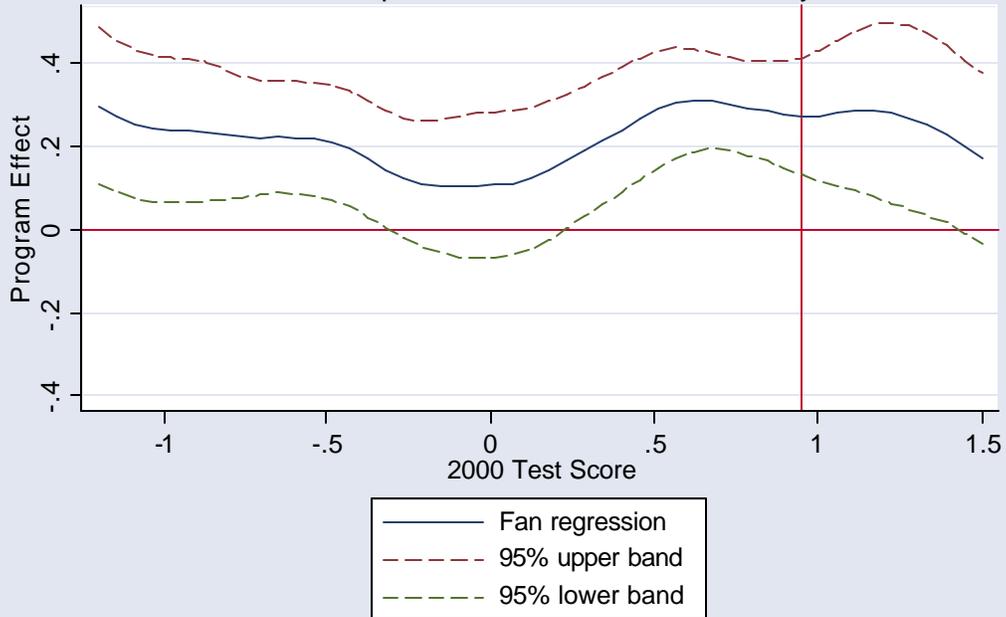
Table 1: Summary Statistics

Panel A: School characteristics	Obs.					
Number of Schools: Program	63					
Number of Schools: Comparison	64					
Number of Schools: Busia	69					
Number of Schools: Teso	58					
Panel B: Student baseline sample	Cohort 1, Baseline			Cohort 2, Baseline		
	Obs	Mean	Std dev	Obs	Mean	Std dev
Number of students: Program	2722			3260		
Number of students: Comparison	2638			3120		
Number of students: Busia district	3162			3761		
Number of students: Teso district	2198			2619		
Gender (1=Male)	5360	0.51	0.50	6380	0.52	0.50
Age in 2001	4937	14.3	1.6	5897	13.3	1.6
Test Score 2000	3217	0.06	0.98	--	--	--
Test Score 2001	4040	0.09	0.99	--	--	--
Test Score 2002	3404	0.05	1.00	3627	0.04	1.03
Mean School Test Score 2000	5286	0.05	0.65	6265	0.07	0.66
Attendance 2001	4805	0.79	0.41	5786	0.77	0.42
Attendance 2002	4787	0.75	0.35	5742	0.75	0.34
Panel C: Student restricted sample	Cohort 1, Restricted			Cohort 2, Restricted		
	Obs	Mean	Std dev	Obs	Mean	Std dev
Number of students: Program	1763			1750		
Number of students: Comparison	1821			1649		
Number of students: Busia district	2366			1847		
Number of students: Teso district	1218			1552		
Gender (1=Male)	3584	0.51	0.50	3399	0.55	0.50
Age in 2001	3584	14.2	1.5	3399	13.12	1.5
Test Score 2000	2347	0.11	0.97	--	--	--
Test Score 2001	3584	0.08	0.99	--	--	--
Test Score 2002	2694	0.10	1.00	3399	0.06	1.00
Mean School Test Score 2000	3584	0.05	0.64	3399	0.06	0.65
Attendance 2001	3584	0.86	0.35	3399	0.84	0.37
Attendance 2002	3454	0.81	0.29	3387	0.87	0.21

Notes: These statistics are for girls and boys in the sample. Dashes indicate that data is currently unavailable (for instance, 2000 and 2001 exams for Cohort 2).

Cohort 1 Baseline sample refers to all students that were registered in grade six in January 2001. Cohort 1 Restricted sample 2001 is restricted to students who were registered in grade six in January 2001, in schools that did not pull out of the program, and for whom we have age, 2001 and 2002 attendance, school average test scores in 2000, and individual test scores in 2001.

Cohort 2 Baseline sample refers to all students that were registered in grade five in January 2001. Cohort 2 Restricted sample 2001 is restricted to students who were registered in grade five in January 2001, in schools that did not pull out of the program, and for whom we have age, 2001 and 2002 attendance, school average test scores in 2000, and test scores in 2002.

Attendance in 2001 is a one-time unannounced visit to schools in term 3, 2001. Attendance in 2002 is consists of three unannounced visits to schools throughout the school year.

Table 2: Demographic and Socio-Economic Characteristics Across Program and Comparison schools in 2002, Cohort 1 and Cohort 2 Girls

	Program	Comparison	Difference (s.e.)
Age in 2001	13.8	13.7	0.1 (0.1)
Gender (1=Male)	0.52	0.52	-0.00 (0.01)
Father's Education (years)	5.1	5.1	-0.0 (0.4)
Mother's Education (years)	4.2	4.4	-0.2 (0.4)
Total Children in Household	6.2	5.7	0.5 (0.4)
Latrine ownership	0.95	0.94	0.01 (0.01)
Iron roof ownership	0.66	0.71	-0.05 (0.03)
Mosquito net ownership	0.31	0.31	0.01 (0.03)

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Standard errors are in parenthesis. Sample includes baseline students in cohort 1 and cohort 2 in 2001 in treatment and comparison schools in Busia and Teso. Data is from 2002 GSP Student Questionnaire. The sample size is 7408 questionnaires.

Table 3: Program Participation Rates and Missing Data, 2001-2002

	Cohort 1					Cohort 2			
	Program	Comparison	Difference	(s.e.)		Program	Comparison	Difference	(s.e.)
Panel A1: Baseline Sample					**Panel A2: Baseline Sample**				
All	2722	2638			All	3260	3120		
Busia	1550	1612			Busia	1846	1915		
Teso	1172	1026			Teso	1414	1205		
Panel B1: Baseline students with 2001 test scores					**Panel B2: Baseline students with 2002 test scores**				
% with 2001 scores, Busia	0.82	0.77	0.04	(0.03)	% with 2002 scores, Busia	0.52	0.51	0.01	(0.04)
% with 2001 scores, Teso	0.63	0.77	-0.14***	(0.04)	% with 2002 scores, Teso	0.64	0.66	-0.02	(0.08)
Panel C1: Baseline students in schools that did not pull out					**Panel C2: Baseline students in schools that did not pull out**				
% in schools that did not pull out, Busia	0.96	1.00	-0.04	(0.04)	% in schools that did not pull out, Busia	0.95	1.00	-0.05	(0.04)
% in schools that did not pull out, Teso	0.88	0.87	0.00	(0.12)	% in schools that did not pull out, Teso	0.88	0.89	-0.01	(0.10)
Panel D1: Restricted Sample: Have 2000 and 2001 Test Scores, Age, and Attendance Data and in schools that did not pull out					**Panel D2: Restricted Sample: Have 2000 and 2002 Test Scores, Age, and Attendance Data and in schools that did not pull out**				
% in Restricted Sample, Busia	0.73	0.76	-0.03	(0.05)	% in Restricted Sample, Busia	0.48	0.50	-0.03	(0.04)
% in Restricted Sample, Teso	0.54	0.58	-0.04	(0.10)	% in Restricted Sample, Teso	0.61	0.57	0.04	(0.09)

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. The relatively low rates of missing data for Teso district students in 2002 is likely the result of the use of 2002 ICS exam scores, rather than district mock exam scores; the 2002 Teso district exams were cancelled due to the upcoming Kenyan national elections (as described in Section 4.2 of the text). Cohort 2 data for Busia district students in 2002 is based on the 2002 Busia district mock exams, which were administered as scheduled in late 2002, and for which students need to pay a small fee (unlike the ICS exams, where were free)..

Table 4: Cohort 1: Baseline Difference Between Attritors and Non-Attritors, Program versus Comparison Schools

	Attritors			Non-Attritors			Difference Attritors – Non-Attritors	
	Program (i)	Comparison (ii)	Difference (iii)	Program (iv)	Comparison (v)	Difference (vi)	Program (i) – (iv)	Comparison (ii) – (v)
Panel A: Attrition from Baseline sample to those in schools that pulled out								
Score in 2000, Busia	0.19	----	----	0.05	0.00	0.05	0.14	----
						(0.19)	(0.13)	
Score in 2000, Teso	0.96	-0.53	1.48**	0.08	0.08	0.00	0.88**	-0.61**
			(0.45)			(0.08)	(0.37)	(0.25)
Panel B: Restricted Sample: Attrition from Baseline sample to those in schools that pulled out and had 2000 and 2001 test scores, age, and 2001 attendance data								
Score in 2000, Busia	0.00	-0.45	0.45**	0.08	0.11	-0.03	-0.07	-0.56***
			(0.19)			(0.19)	(0.11)	(0.08)
Score in 2000, Teso	0.18	-0.20	0.37	0.20	0.11	0.09	-0.02	-0.30
			(0.30)			(0.18)	(0.23)	(0.20)

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Sample attrition involves those baseline students who are not included in the restricted sample, that is all schools that voluntarily pulled out of the program or students whose baseline attendance, age, and 2000, 2001, or 2002 test score data was not available. Dashes (----) indicate that there were no attritors.

Table 5: Program Impact on Test Scores

	Dependent variable: Normalized test scores from 2001 and 2002
Panel A: Restricted Sample	Girls and Boys, Busia and Teso districts OLS
Program school	0.11**
	(0.05)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes
Sample Size	10072
R^2	0.30
Mean of dependent variable	0.08

	Girls and Boys, Busia district
Panel B: Restricted Sample	OLS
Program school	0.19***
	(0.07)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes
Sample Size	6168
R^2	0.36
Mean of dependent variable	0.09

	Girls, Busia district	Boys, Busia district
Panel C: Restricted Sample	OLS	OLS
	(1)	(2)
Program impact, Cohort 1 (in 2001)	0.27***	0.20*
	(0.11)	(0.10)
Program impact, Cohort 2 (in 2002)	0.22**	0.13
	(0.10)	(0.12)
Post-Program impact, Cohort 1 (in 2002)	0.24***	0.13
	(0.08)	(0.09)
Mean school test score, 2000	0.81***	0.78***
	(0.07)	(0.06)
Student age in program year	-0.02***	-0.03***
	(0.01)	(0.01)
% Luo ethnic group in school	0.11	0.11
	(0.20)	(0.25)
% Teso ethnic group in school	0.54**	0.29
	(0.26)	(0.26)
% other (non-Luhya, Luo, Teso) ethnic group in school	-0.11	1.00**
	(0.52)	(0.43)
Sample Size	2931	3237
R^2	0.40	0.37
Mean of dependent variable	-0.04	0.19

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized to make the mean of comparison test scores zero and standard deviation equal to one. Indicator variables are included for Cohort 1 in 2001, Cohort 1 in 2002, and Cohort 2 in 2002 (coefficient estimates not shown). Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Busia district.

Table 6: Program Impact on Test Scores by Subject,
Busia District Girls (Restricted Sample)

	Math OLS (1)	Science OLS (2)	English OLS (3)	Swahili OLS (4)
Program impact, Cohort 1 (in 2001)	0.24	0.26**	0.25***	0.09
	(0.16)	(0.12)	(0.09)	(0.09)
Program impact, Cohort 2 (in 2002)	0.29**	0.28***	0.07	0.07
	(0.12)	(0.10)	(0.10)	(0.10)
Post-Program impact, Cohort 1 (in 2002)	0.23**	0.24**	0.17***	0.25***
	(0.11)	(0.10)	(0.07)	(0.10)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes	Yes	Yes
Sample Size	2931	2931	2931	2931
R^2	0.18	0.24	0.37	0.22
Mean of dependent variable	-0.02	-0.12	0.02	0.11
F-test, impacts equal across cohorts (p-value)	0.89	0.91	0.24	0.24

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized to make the mean of comparison test scores zero and standard deviation equal to one. School ethnic composition controls include % Luo ethnic group in school, % Teso ethnic group in school, and % other (non-Luhya, Luo, Teso) ethnic group in school. Indicator variables are included for Cohort 1 in 2001, Cohort 1 in 2002, and Cohort 2 in 2002 (coefficient estimates not shown). Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Busia district.

Table 7: Program Impact on School Attendance, Busia District

Panel A: Overall student attendance in 2001 and 2002, Busia District (Restricted Sample)	Dependant variable: Attendance in 2001, 2002
	Boys and Girls, Busia district
	OLS
Program school	0.05**
	(0.02)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes
Sample Size	8362
R^2	0.03
Mean of dependent variable	0.84

Panel B: Student attendance 2001 and 2002, Busia District (Restricted Sample)	Dependent variable: Attendance in 2001, 2002	
	Girls, Busia district	Boys, Busia district
	OLS	OLS
	(1)	(2)
Program impact, Cohort 1 (in 2001)	0.06	0.08*
	(0.04)	(0.05)
Program impact, Cohort 2 (in 2002)	0.01	-0.03
	(0.02)	(0.02)
Post-Program impact, Cohort 1 (in 2002)	0.02	0.02
	(0.02)	(0.03)
Pre-Program impact, Cohort 2 (in 2001)	0.10**	0.10*
	(0.05)	(0.06)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	3994	4368
R^2	0.89	0.87
Mean of dependent variable	0.86	0.83

Panel C: Teacher attendance in 2002, Busia District	Dependent variable: Teacher attendance in 2002, Busia district
	OLS
Program school	0.05***
	(0.02)
Mean school test score in 2000, and school ethnic composition controls	Yes
Sample Size	2399
R^2	0.02
Mean of dependent variable	0.86

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. School ethnic composition controls include % Luo ethnic group in school, % Teso ethnic group in school, and % other (non-Luhya, Luo, Teso) ethnic group in school. Indicator variables are included for Cohort 1 in 2001, Cohort 1 in 2002, Cohort 2 in 2001, and Cohort 2 in 2002 in Panel A (coefficient estimates not shown).

Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Busia district. Results are similar for the unrestricted sample (not shown). The 2001, 2002 school attendance measure takes on a value of one if the student was present in school on the day of an unannounced attendance check. There was one such student attendance observation in the 2001 school year, and three in 2002. The teacher attendance visits were also unannounced, and actual teacher presence at school recorded. Teachers included Head Teachers, Deputy Head Teachers, and all grades, in both program and comparison schools.

Table 8: Program Impact on Education Habits, Inputs, and Attitudes in 2002, Busia District Girls (Restricted Sample)

	Cohort 1:		Cohort 2:	
	Estimated program impact	Mean of dep. var.	Estimated program impact	Mean of dep. var.
Dependent Variables:				
Panel A: Study habits				
Student used a textbook at home in last week	0.06**	0.72	-0.03	0.72
	(0.03)		(0.04)	
Student did homework in last two days	0.01	0.79	0.04	0.79
	(0.05)		(0.05)	
Student handed in homework in last week	0.00	0.88	0.01	0.87
	(0.04)		(0.05)	
Student went for extra coaching in last two days	0.00	0.84	-0.01	0.79
	(0.05)		(0.05)	
Teacher asked the student a question in class in last two days	-0.03	0.36	0.07	0.35
	(0.04)		(0.05)	
Panel B: Educational Inputs				
Number of textbooks at home	0.01	1.9	0.14	1.8
	(0.18)		(0.13)	
Number of new textbooks this term	0.05	2.8	0.08	3.0
	(0.15)		(0.10)	
Number of new exercise books this term	0.11	0.9	0.18	0.9
	(0.16)		(0.12)	
Panel C: Attitudes towards education				
Student thinks she a "good student"	0.02	0.69	0.00	0.75
	(0.06)		(0.05)	
Student thinks that being a "good student" means "working hard"	-0.03	0.84	-0.02	0.76
	(0.03)		(0.03)	
Student thinks can be in top three in her class	0.03	0.37	-0.01	0.35
	(0.05)		(0.05)	
Student prefers school to other activities (index)	0.01	0.70	0.02	0.72
	(0.01)		(0.02)	

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Marginal probit coefficient estimates are presented when the dependent variable is an indicator variable, and OLS regression is performed otherwise. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Each coefficient estimate is the product of a separate regression, where the explanatory variables are an indicator for being in a program school, as well as student age, mean school test score in 2000, and school ethnic composition controls. School ethnic composition controls include % Luo ethnic group in school, % Teso ethnic group in school, and % other (non-Luhya, Luo, Teso) ethnic group in school (analogous to Table 5, Panel C, regression 1).

Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Busia district. Results are similar for the unrestricted sample (not shown). The sample size varies from approximately 700-850 observations, depending on the extent of missing data on the dependent variable.

The "student prefers school to other activities" index is the average of eight binary variables indicating whether the student prefers a school activity (coded as 1) or a non-school activity (coded 0). The school activities include: doing homework, going to school early in the morning, and staying in class for extra coaching. The non-school activities include fetching water, playing games or sports, looking after livestock, cooking meals, cleaning the house, or doing work on the farm.

Table 9: Program Impacts on Self-Esteem and Test Scores in 2002, Busia District Girls (Cohort 1 Restricted Sample), Regression Discontinuity Estimates

	Dependent variable:		
	Student thinks she is a "good student"	Student prefers school to other activities	2002 test score
	Probit	Probit	OLS
	(1)	(2)	(3)
Program school *	0.09	-0.02	-0.08
2001 test score above winning threshold	(0.14)	(0.07)	(0.19)
Program school	0.03	0.02	0.05
	(0.07)	(0.02)	(0.11)
2001 test score above winning threshold	-0.03	0.01	-0.01
	(0.12)	(0.06)	(0.16)
2001 test score polynomial controls	Yes	Yes	Yes
Student age, school ethnic composition controls	Yes	Yes	Yes
Sample Size	707	707	707
R^2	0.02	0.04	0.58
Mean of dependent variable	0.69	0.69	-0.02

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Marginal probit coefficient estimates are presented. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized to make the mean of comparison test scores zero and standard deviation equal to one. The 2001 test score winning threshold in Busia district was 0.95 s.d.

The 2001 test score polynomial controls include linear, squared, and cubic 2001 test score terms, and these three terms interacted with the program school indicator variable. School ethnic composition controls include % Luo ethnic group in school, % Teso ethnic group in school, and % other (non-Luhya, Luo, Teso) ethnic group in school. Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Busia district. In Table 9, the sample is further restricted to those students with complete data for all three dependent variables.

The "student prefers school to other activities" index is the average of eight binary variables indicating whether the student prefers a school activity (coded as 1) or a non-school activity (coded 0). The school activities include: doing homework, going to school early in the morning, and staying in class for extra coaching. The non-school activities include fetching water, playing games or sports, looking after livestock, cooking meals, cleaning the house, or doing work on the farm.

Appendix Table A: Timeline of the Girls Scholarship Program 2001-2002

Time		Activity
November	2000	Fifth grade students in cohort 1 take exams that will be used as baseline scores for analysis.
March	2001	Announced Girls' Scholarship Program to Head Teachers in all treatment schools. Head Teachers disseminate information to parents and students.
June	2001	Lightning strikes school in Teso
September – October	2001	NGO holds Parent-Teacher meetings in all schools to remind parents and students of the program and upcoming tests.
September – October	2001	Field officers perform unannounced visits to schools to collect attendance data.
November	2001	Sixth grade students in cohort 1 take exams that will be used to measure impact of the program.
January	2002	NGO holds school assemblies to announce winners and give scholarships.
January – March	2002	Field officers perform unannounced visits to schools to collect attendance data.
February – June	2002	Field officers administer student survey to all fifth, sixth and seventh graders.
November	2002	End of year exams in Busia. Sixth grade Busia students in cohort 2 take exams that will be used to determine second cohort of scholarship winners. Seventh grade Busia students in cohort 1 take exams that will be used to determine second year impacts of the scholarship program. Teso exams canceled due to National elections.
February	2003	NGO administers standardized exams in both Busia and Teso. Sixth grade Teso students in cohort 2 take exams that will be used to determine second cohort of scholarship winners. Seventh grade Teso students in cohort 1 take exams that will be used to determine second year impacts of the scholarship program.

Appendix Table B: Program Impact on Test Scores, Teso District

	Dependent variable: Normalized test scores from 2001 and 2002	
Panel A: Restricted Sample	Girls and Boys, Teso District OLS	
Overall program impact	-0.03	
	(0.07)	
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	
Sample Size	4000	
R^2	0.19	
Mean of dependent variable	0.02	

Panel B: Restricted Sample	Girls, Teso district OLS (1)	Boys, Teso district OLS (2)
Program impact, Cohort 1 (in 2001)	-0.02	-0.06
	(0.10)	(0.10)
Program impact, Cohort 2 (in 2002)	0.01	-0.05
	(0.10)	(0.13)
Post-Program impact, Cohort 1 (in 2002)	-0.04	-0.03
	(0.11)	(0.08)
Mean school test score, 2000	0.66***	0.68***
	(0.09)	(0.09)
Student age in program year	-0.08***	-0.10***
	(0.02)	(0.02)
% Luo ethnic group in school	1.62	-0.36
	(1.70)	(1.67)
% Teso ethnic group in school	0.30	-0.29
	(0.46)	(0.53)
% other (non-Luhya, Luo, Teso) ethnic group in school	0.16	-0.34
	(0.73)	(0.64)
Sample Size	1840	2160
R^2	0.24	0.20
Mean of dependent variable	-0.13	0.15

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Huber robust standard errors are in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized to make the mean of comparison test scores zero and standard deviation equal to one. School ethnic composition controls include % Luo ethnic group in school, % Teso ethnic group in school, and % other (non-Luhya, Luo, Teso) ethnic group in school. Indicator variables are included for Cohort 1 in 2001, Cohort 1 in 2002, and Cohort 2 in 2002 (coefficient estimates not shown). Restricted sample includes students who were registered in sixth grade in January 2001, did not pull out of the program, and for whom we have age, 2001 attendance, and test scores in 2000 and 2001 in Teso district.