Incentives to Learn

Michael Kremer*

Edward Miguel**

Rebecca Thornton***

September 2004

Abstract: We report results from a randomized evaluation of a merit scholarship program for adolescent girls in Kenya. Girls who scored well on academic exams had their school fees paid and received a cash grant for school supplies. Girls eligible for the scholarship showed significant gains in academic exam scores (average gain 0.17 standard deviations) and these gains persisted following the competition. There is also evidence of positive program externalities on learning: boys, who were ineligible for the awards, also showed sizeable average test gains, as did girls with low pretest scores, who were unlikely to win. Both student and teacher school attendance increased in the program schools.

* Dept. of Economics, Harvard University, The Brookings Institution, and NBER. Littauer 207, Harvard University, Cambridge, MA 02138, USA; mkremer@fas.harvard.edu.
** Dept. of Economics, University of California, Berkeley and NBER. 549 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, USA; emiguel@econ.berkeley.edu.
*** Dept. of Economics, Harvard University, Littauer 207, Cambridge, MA 02138, USA; rlthornt@fas.harvard.edu.

The authors thank ICS Africa and the Kenya Ministry of Education for their cooperation in all stages of the project, and would especially like to acknowledge the contributions of Elizabeth Beasley, Pascaline Dupas, James Habyarimana, Sylvie Moulin, Robert Namunyu, Petia Topolova, Peter Wafula Nasokho, Owen Ozier, Maureen Wechuli, and the GSP field staff and data group, without whom the project would not have been possible. George Akerlof, David Card, Rachel Glennerster, Brian Jacob, Matthew Jukes, Victor Lavy, Michael Mills, Antonio Rangel, Joel Sobel, Doug Staiger, and many seminar participants have provided valuable comments. We are grateful for financial support from the World Bank and MacArthur Foundation. All errors are our own.

1. Introduction

Many scholarships in the United States were merit-based historically, but during the 1960s and 1970s there was a dramatic move toward need-based awards. Recently, however, there has been a resurgence in merit scholarships: while more than three-quarters of all state-funded college scholarships in the United States are now based on financial need, merit funds have grown by almost 50% in the past five years (College Board 2002). Merit scholarships are potentially attractive because they may help channel educational investments to those with the highest return – if education and initial achievement are complements, and if credit constraints sometimes prevent those with high returns to education from obtaining it – and they may also provide increased incentives for study effort.

Understanding the impact of study effort on educational outcomes is valuable in its own right. While most education research focuses on the effect of material inputs, class size, or school organization, the most important input in the education production function may be study effort. Study effort may be systematically suboptimal because many adolescents have time inconsistent preferences, such as hyperbolic discounting, or higher time discount rates than adults (Greene 1986, Nurmi 1991), or because students do not fully capture the benefits of their effort due to human capital externalities,

This paper examines the impact of a merit scholarship program introduced in rural Kenyan primary schools in 2001, which provided awards 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$360 (World Bank 2002). Schools in which the scholarship program was introduced were randomly selected from a group of candidate schools, allowing us to attribute differences in educational outcomes between the program and comparison schools to the program.

We find that girls in the program schools had significantly higher test scores than those in comparison schools. Moreover, there is evidence the program generated substantial positive classroom externalities: in the larger of the two study districts (Busia district) there were significant test gains for boys (boys were all ineligible for the scholarship), as well as for girls with low pretest scores, who had little chance of winning.

Many argue that private incentives to invest in education are too weak because education generates externalities (Lucas 1988), and such externalities are often cited as a justification for government education subsidies. However, empirical studies suggest human capital externalities are small, if they exist at all (Acemoglu and Angrist 2000, Moretti 2004). All these studies examine positive externalities in the labor market, but our results suggest it may well be that the largest positive externalities from education occur earlier, within the classroom (Lazear 2001).

The program was implemented in two Kenyan districts, Busia and Teso. In Busia, the larger district, the program was received well, but in Teso it was not, and there was substantial attrition from the program. Point estimates that do not correct for attrition suggest a very strong program impact on test scores in Busia and insignificant impacts in Teso. However, it is unclear whether this reflects a smaller program effect or the high and asymmetric attrition rates in Teso district.

In terms of the underlying behavioral channels, student school attendance was significantly higher in Busia district program schools, evidence that study effort increased there in response to the incentive. School attendance increased for both girls and boys in Busia program schools, and this apparent strategic complementarity in student effort suggests that small changes in exogenous factors could lead to large changes in effort, and perhaps even multiple equilibria in educational outcomes. Girls in program schools were also somewhat more likely to use textbooks to study at home, further evidence that student effort increased in program schools. There is some suggestive, though ultimately inconclusive, evidence of increased parental inputs into education, proxied by the purchase of additional textbooks and exercise books for children.

The program increased teacher attendance in Busia but not in Teso, and one plausible explanation is that parents in Busia placed greater pressure on teachers to improve their performance. Community enthusiasm for the program may also have increased the non-monetary utility benefits of winning the award in Busia, in terms of local social prestige, for instance, motivating both students and teachers to exert additional effort.

There is no evidence that the scholarship program simply led students to focus on their test performance at the expense of other dimensions of learning. This stands in sharp contrast to another project conducted by the same non-governmental organization which provided incentives for teachers based on students' test scores. That teacher incentive program had no measurable effect on either student or teacher attendance, but increased the frequency of test preparation sessions (Glewwe et al. 2003). Students' scores increased on the exam for which the teacher incentives were provided, but did not remain high afterwards. In contrast, in the girls' scholarship program we study, both student and teacher attendance increased, and test score gains remain large in the year following the competition.

There is no evidence (from surveys of students) that program incentives weakened the intrinsic motivation to learn in school. There are no statistically significant changes in students' self-expressed attitudes toward school, or toward their own academic ability, or in students' time use outside of school. While standard economic models suggest incentives should increase individual study effort, an alternative theory from psychology asserts that extrinsic rewards may interfere with intrinsic motivation and actually reduce effort.[1] A weaker version of this view is that incentives lead to better performance in the short-run, but have negative effects after the incentive is removed by weakening intrinsic motivation, but we find no evidence of this when we examine test scores in the year following the scholarship competition (or at least any reduction in intrinsic motivation was offset by other factors).

In the work most closely related to the current study, Angrist and Lavy (2002) find that cash awards raised test performance among 500 high school students in Israel. They examine a pilot scholarship program that provided cash for good performance on matriculation exams in twenty schools. Students offered the merit award were approximately 6-8 percentage points more likely to successfully

[1] Early experimental psychology research in education supported the idea that reward-based incentives lead to increased effort in students (Skinner 1961). However, laboratory research conducted in the 1970's studied behavior before and after individuals received "extrinsic" motivational rewards, and found that these external rewards produced negative impacts in some situations (Deci 1971; Kruglanski et al. 1971; Lepper et al. 1973). Later laboratory research attempting to quantify the effect of external factors on intrinsic motivation has yielded mixed conclusions: Cameron et al. (2001) conducted meta-studies of over 100 experiments and found that the negative effects of external rewards were limited and could be overcome in certain settings – such as for high-interest tasks – but in a similar meta-study Deci et al. (1999) conclude that there are usually negative effects of rewards on task interest and satisfaction. The current study differs from much of the existing work by estimating impacts in a real-world context rather than the laboratory, and by exploring spillover effects on third parties.

pass their exams than comparison students in a pilot program that randomized awards among schools, with the largest effects among the top quartile of students. A smaller pilot that randomized awards at the individual level within a different set of schools did not produce significant impacts.

This program differs from ours in several important ways. First, due to political and logistical issues, the program in Israel and its evaluation, which was meant to run for three years, were discontinued after the first year – making it impossible to estimate longer-term impacts, and impacts once the incentive was removed. Second, the sample in the current study includes more than three times as many schools as their pilot study The sample of students in the Angrist and Lavy study was not large enough to ensure that average characteristics in the randomly assigned program and comparison groups of schools were similar. Third, in addition to test score outcomes, we collected data on student school attendance, teacher attendance, purchases of school supplies, student time use, and a range of student attitudes which allow us to explore the mechanisms through which merit scholarships affect learning, unlike Angrist and Lavy, who do not have such data. Nor are they able to estimate externality impacts of increased student effort – although note that the large estimated impacts in the pilot that randomized incentives across schools, relative to the pilot that randomized incentives across individuals within the same school, is consistent with the existence of within classroom externalities in student effort.[2]

A number of studies suggest university scholarships increase enrollment (for instance, Dynarski 2003) though the few studies that examine the incentive effects of merit scholarships find mixed impacts. Binder et al. (2002) find that while scholarship eligibility in New Mexico increased student grades, the

[2] Leuven et al (2003) also use an experimental design, to estimate the effect of a financial incentive on the performance of Dutch university students, but their small sample size limits statistical precision, complicating inference. Ashworth et al. (2001) study Education Maintenance Allowances (EMA), weekly allowances given to 16-19 year old students from low-income U.K. households based on school enrollment and academic achievement. Initial findings indicate that EMA raised school enrollment among eligible youth by 5.9 percentage points and by 3.7 percentage points among the ineligible, suggesting externalities. It is unclear how much of these impacts are due to rewarding students for enrollment versus achievement. Since program areas were not randomly selected – EMA was targeted to poor urban areas – the authors resort to propensity score matching to estimate impacts. Croxford et. al. (2002) find similar EMA impacts in Scotland. Angrist et al (2002) show that a Colombian program that provided vouchers for private schools to students conditional on their maintaining a satisfactory level of academic performance led to academic gains, although it is unclear how much of this impact came from the expanded range of school choice participants experienced, and how much from the incentive.

number of credit-hours students completed decreased – suggesting that students took fewer courses in order to keep up their grades. Similarly, after the HOPE college scholarship program was introduced the average SAT score for Georgia's high school seniors rose almost 40 points (Cornwell et al. 2002), but it resulted in a 2% average reduction in completed college credits, 12% decrease in full course-load completion, and 22% increase in summer school enrollment (Cornwell et al 2003), presumably again to boost grades, thus undermining the key program objective of increased learning. But these potential distortions are not relevant in the setting we examine where the curriculum is fixed.

The paper proceeds as follows. Section 2 provides information on schooling in Kenya and on the scholarship program. Section 3 presents a model of incentives and study effort. Section 4 discusses the data, section 5 presents the estimation strategy and results, and section 6 compares the cost effectiveness of merit scholarships to other programs. The final section concludes.

2. The Girls Scholarship Program

2.1 Schooling in Kenya

Schooling in Kenya consists of eight years of primary school followed by four years of secondary school. While most children enroll in primary school – approximately 85% of children of primary school age in western Kenya are enrolled in school (Central Bureau of Statistics 1999) – there are high dropout rates in grades 5, 6, and 7, about one-third finish primary school, and only a fraction of these students enter secondary school. The dropout rate is especially high for teenage girls.[3] Admission to secondary school depends on performance on the government Kenya Certificate of Primary Education (KCPE) exam in Grade 8, and students take that exam quite seriously. To prepare for the KCPE, students in grades 4-8 typically take standardized exams at the end of each school year – although these exams are sometimes canceled, for example, due to teacher strikes or fears of election year violence. End-of-year exams are standardized for each district and test students in five subjects: English, geography/history, mathematics, science, and Swahili. Students must pay a fee to take the exam, US$1-2 depending on the year, and we

[3] For instance, girls in our baseline sample (in comparison schools) had a dropout rate of 9% from January 2001 through early 2002, versus 6% for boys. Drop-out rates were slightly lower in program schools (not shown).

discuss implications of this fee below. Kenyan district education offices have a well-established system of exam supervision, with proctors (called "invigilators") from outside the school monitoring all exams, and teachers from that school playing no role in either exam supervision or grading. Invigilators document and punish all instances of cheating, and report these cases back to the district education office.

The Girls Scholarship Program (GSP) was carried out by a Dutch non-governmental organization (NGO), called ICS Africa, in two rural districts in western Kenya, Busia and Teso. Busia district is mainly populated by a Bantu-speaking ethnic group (the Luhya) with agricultural traditions while Teso district is populated primarily by a Nilotic-speaking ethnic group (the Teso) with pastoralist traditions. These groups differ in language, history, and certain present-day customs, although not typically along observed household assets. The two districts were originally part of a single district which was partitioned in 1995. ICS Africa is headquartered in Busia district, and most of its staff (including those who worked on the scholarship project) are ethnic Luhyas.

Speaking in broad terms, a common perception in western Kenya is that the Teso community is less "progressive" than the Luhya community. Historically, Tesos in this area were educationally disadvantaged relative to Luhyas, with fewer Teso than Luhya secondary school graduates, for example. Project survey data (described below) confirms this disparity between the districts: parents of students in Teso district have 0.4 years less schooling than Busia district parents on average. There is a tradition of suspicion of outsiders in Teso district, and this has at times led to misunderstandings between NGO's and some people there. It has also been claimed that indigenous religious beliefs, traditional taboos and witchcraft practices remain stronger in Teso than in Busia (Government of Kenya 1986).

When the scholarship program was introduced primary school in the area charged school fees to cover non-teacher costs including textbooks, chalk, and classroom repair. These fees averaged approximately US$6.40 (500 KSh)[4] per family each year. In practice, while these fees set a benchmark for bargaining between parents and headmasters, most parents did not pay the full fee. In addition to this per family fee, there were also fees for particular activities, such as taking standardized exams (noted

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

above), and families had to pay for their children's school supplies, exercise books, certain textbooks, and uniforms (the average uniform costs US$6.40).

2.2 Project Description and Timeline

Half of the sample of 127 sample primary schools were randomly invited to participate in the Girls Scholarship Program in March 2001. The randomization was stratified by administrative divisions (there are eight divisions in Busia and Teso districts), and by participation in a past NGO assistance program, which had provided classroom flip charts to some schools.[5] Randomization was done using a computer random number generator, and as we discuss below (Section 4), this procedure was successful at creating program and comparison groups largely similar along observable characteristics.

The program provided incentives for students to excel on academic exams, beyond the usual benefits to good academic performance. The scholarship program provided winning Grade 6 girls with an award for the next two academic years, Grades 7 and 8 (through the end of primary school – the selection of winners is described below). In each year, the award consisted of: (1) a grant of US$6.40 (500 KSh) intended to cover the winner's school fees and paid directly to her school; (2) a grant of US$12.80 (1000 Kenyan shillings, KSh) paid to the girl's family and intended for school supplies; and (3) public recognition at a school awards assembly organized by the NGO.[6]

Given that many parents would not otherwise have fully paid school fees, primary schools with winners benefited to some degree from the award money that paid winners' fees.[7] Some of these funds may have also benefited teachers, if they were used to improve the staff room or pay for refreshments for teachers, for instance, although the amounts involved in this were likely small.

---

[5] All GSP schools had previously participated in an evaluation of a flip chart program, and are a subset of that sample. The flip chart evaluation schools were chosen since they had not been recipients of previous NGO school assistance programs and were also not relatively well-off. Half of the schools in that evaluation received flip charts and half did not. There is no evidence that the flip chart program affected test scores. These schools are representative of local primary schools along most dimensions – see Glewwe et al. (2004) for details on the sample.

[6] Note that there may be benefits for winners' siblings from the income transfer because: (i) primary school fees were levied per household rather than per student, so the cost of schooling declined for siblings as well, and (ii) potential within household learning spillovers. We plan to estimate sibling impacts in future research.

[7] Although mandatory school fees were abolished in early 2003, as described above, the NGO continued to pay grant money directly to schools with scholarship winners in 2003 and 2004.

In the two years of the program, two cohorts of Grade 6 girls competed for scholarships. Girls registered for Grade 6 in January 2001 in program schools were the first eligible cohort (cohort 1) and those registered for grade 5 in January 2001 made up the second cohort (cohort 2), and they competed for the award in 2002. The NGO restricted eligibility to girls who were already enrolled in a program school in January 2001, before the program was announced. Thus there was no incentive for students to transfer into program schools, and incoming student transfer rates were low and nearly identical in program and comparison schools (not shown).

In November 2000, cohort 1 students took end-of-year Grade 5 exams, and these are used as baseline test scores in the evaluation.[8] In March 2001, the NGO held meetings with the headmasters of schools invited to participate in the program to inform them of program plans and to give each school community the choice to participate. Headmasters were asked to relay information about the program to parents via a school assembly. Because of variation in the extent to which headmasters effectively disseminated this information, there was a sense that awareness was inadequate in some areas, and as a result the NGO held additional community meetings in September and October to reinforce knowledge about program rules in advance of the November 2001 district exams. After the meetings, enumerators began collecting school attendance data during unannounced visits.

Students took district exams in November 2001, and each district gave a separate exam. Scholarship winners in Grade 6 were chosen based on their total score across all five subject tests. The NGO then awarded scholarships to the highest scoring 15% of Grade 6 girls in the program schools in each district (this amounted to 110 girls in Busia district and 90 in Teso). Schools varied considerably in the number of winners, but 57% of program schools (36 of 63 schools) had at least one 2001 winner; among schools with at least one 2001 winner, there was an average of 5.6 winners per school.

Scholarship winners differ from non-winners in certain family background dimensions. Most importantly, average years of parent education is nearly three years greater for scholarship winners than

[8] A detailed project timeline is presented in Appendix Table A. Unfortunately, there is incomplete 2000 baseline exam data for cohort 2 (when these students were in grade 4), and thus baseline comparisons focus on cohort 1. Average 2000 scores are used to control for baseline differences across schools, as described below.

losers (7.7 years for winners versus 4.8 years for non-winners), and this large effect is statistically significant at 99% confidence. However, there is no statistically significant difference between winners and non-winners in terms of household ownership of iron roofs or latrines (not shown), suggesting that children from wealthier households in terms of asset ownership were no more likely to win (though this remains somewhat speculative in the absence of detailed household consumption expenditure data).[9]

The NGO held school assemblies – for students, parents, teachers, and local government officials – in January 2002 to announce and publicly recognize the 2001 winners. Each winner was awarded a certificate, parents received the US$12.80 (1000 Ksh) cash grant, and the school received US$6.40 (500 Ksh) to cover the winner's school fees. The community was reminded that the program would continue for one more year. Parents of the winning girls were instructed that the grant should be used to purchase school-related materials for the winning girl, such as a school uniform, textbooks, and exercise books.[10]

During the 2002 academic year, the NGO returned to both program and comparison schools to conduct unannounced attendance checks and administer questionnaires to students in Grades 5-7. These surveys collect information on study effort, habits, and attitudes toward school (described below). District exams were again held in late 2002 in Busia district. Primary school exams in Teso district were canceled in 2002 because of possible disruptions in the run-up to upcoming 2002 national elections and a threatened teacher strike, so the NGO instead administered standardized academic exams in February 2003 there. Thus the second cohort of scholarship winners were chosen in Busia district based on the official 2002 district exam, while Teso district winners were chosen based on the NGO exam. In this second round of the scholarship competition, 70% of the program schools (44 of 63 schools) had at least one winner, an increase over 2001.

[9] When the top 15% of cohort 1 girls on the 2001 exams are contrasted in the program versus comparison groups (where the top 15% measure is constructed separately for the two groups), there are no statistically significant differences in the predictive power of household socioeconomic characteristics across the two treatment groups in Busia district (where the dependent variable is the top 15% indicator – regressions not shown).

[10] Structured interviews with several teachers and winning girls indicated that the award money, at least in part, did in fact often go towards purchasing items such as books, uniforms, math sets, and watches for the winner. However, this is impossible to test formally without detailed household consumption expenditure data, which we do not have.

The student survey data indicates that most girls understood program rules by 2002: 89% of cohort 1 and 2 girls in Busia district claimed to have heard of the program, and knowledge levels were only slightly lower in Teso district (86%). Girls had somewhat better knowledge about program rules governing eligibility and winning than boys: Busia girls were 7 percentage points more likely than boys to know that “only girls are eligible for the scholarship” (86% for girls versus 79% for boys), although the proportion among boys is still very high, suggesting that the vast majority of boys knew that they were ineligible; patterns are again similar in Teso district (not shown). Note that random measurement error is likely to be reasonably large for these student surveys, since rather than being filled in by an enumerator who individually interviews students, the surveys were filled in by students (at their desks) with the enumerator explaining the questionnaire to the class as a whole. Thus values of 100% are unlikely even if all students had excellent program knowledge. Girls were very likely (70%) to report that their parents had mentioned the program to them, suggesting some parental involvement.

In the run-up to the 2002 national elections, in late 2001 then-president Daniel arap Moi announced a national ban on primary school fees, but the central government did not provide alternative sources of funding to schools and other policymakers were unclear on whether schools could impose “voluntary fees” to cover school inputs. As a result, school committees in this area generally continued collecting some fees in 2002, but fund-raising appears to have fallen somewhat (although we do not have quantitative evidence on the extent of the decline). Mwai Kibaki became president of Kenya following December 2002 elections, and eliminated primary school fees in early 2003. This policy was quickly implemented by almost all local school committees – in part because the national government made substitute payments to schools to replace local fees, financed by a World Bank loan. This national policy change with regards to fees came into effect after the study period of March 2001 to February 2003, and is unlikely to have affected our results. The NGO preserved the program design after this policy change, and in particular awards for winners’ families and schools were made in 2003 and 2004.

In June 2001, lightning struck a primary school in Teso district (Korisai Primary School, not in the GSP sample), severely damaging the school, killing seven students, and injuring 27 others. Because

ICS had been involved with another assistance program in that school, and due to strange coincidences – for instance, the names of certain lightning victims were the same as the names of ICS staff members who had recently visited the school – the deaths were associated with ICS in the eyes of some community members, and the incident led several schools to pull out of the Girls Scholarship Program: of the original 58 sample schools in Teso district, five pulled out of the program at that time, and one Busia school located near the Teso district border also pulled out. Figure 1 presents the location of the lightning strike and of the schools that pulled out of the program, several of which are located near the lightning strike. Three of the six schools that pulled out of the program were treatment schools, and three were comparison schools. We discuss implications for econometric inference in Section 4.2 below.

Structured interviews were conducted during June 2003 with 64 teachers in 18 program schools, and these suggest there were stark differences in program reception across Busia and Teso districts – perhaps in part due to the lightning strike. The teachers were asked to rate parental support for the program and while 90% of the teachers in Busia claimed that parents were either “Very positive” or “Somewhat positive” toward the scholarship program, the analogous rate in Teso district was only 58%, and this difference across the districts is statistically significant at 99% confidence.

3. Incentives, Externalities, and Study Effort

A stylized framework helps to illustrate how merit awards could impact academic test scores. Individual study effort may take various forms, including some which are relatively easy to observe, such as attending school, and others that are more difficult to measure, such as paying more attention in class. In addition to individual study effort, academic performance may also be a function of: the study effort of other students in the class, since it may be easier to learn when classmates are also studious, a theoretical point developed in Lazear (2001); teacher effort; as well as the child’s current academic ability (or “human capital”), which is a function of the past effort exerted by the child herself, by her classmates, and her teacher, as well as a function of the child’s innate ability. We ignore other inputs into educational production (e.g., textbooks and chalk) in the discussion below for simplicity.

Theoretically, the effort of children and their classmates, and of children and their teachers could potentially be either complements or substitutes. Similarly, own effort and current academic ability may be either complements or substitutes, and thus own effort at one point in time may complement or substitute effort at other times (working through the academic ability term).[11]

Yet it seems plausible that own effort, effort of other students, and teacher effort may be complements. In this case, programs which increase effort by some students could generate multiplier effects in individual effort, and also open up the theoretical possibility of multiple classroom equilibria, some equilibria with high levels of effort by students and others with a poor overall learning environment.[12] Educators often stress the importance of classroom culture and Akerlof and Kranton (2003) have recently attempted to formally model these cultures. The available empirical evidence is also consistent with the existence of multiple equilibria in classroom culture. Most studies find that conventional educational variables – including the pupil-teacher ratio and expenditures on standard inputs, like textbooks – explain only a modest fraction of variation in test score performance, typically with R^2 values on the order of 0.2-0.3 (Summers and Wolfe 1977, Hanushek 2003). While there are many possible interpretations of this finding, one possibility is that unobserved classroom culture is driving much of the test score variation. In the current study, the divergence in educational outcomes and program impacts between Busia and Teso districts (described below), two areas with different local ethnic compositions and traditions, is also consistent with multiple equilibria in classroom culture.

The Girls Scholarship Project that we study directly affected incentives to exert study effort, and this effort increases the probability that an individual will win the scholarship. Winning the scholarship has some value to students, and this value could differ by school due to variation in local non-monetary benefits, such as social prestige from winning. The probability of winning a scholarship is a function of both the individual's test score and assignment to a program school, which takes on a value of one for

[11] Note that other possible channels for persistent effects of the program are the cash grant payment to winners, and the payment of school fees to winners' schools.

[12] Although Cooper and John (1988) restrict attention to multiple symmetric Nash equilibria, unlike the framework here, the main insights of their model are likely to carry over to this setting under certain conditions.

program (“treatment”) schools. The probability of winning the scholarship is zero for all non-eligible students (those in comparison schools, boys, and girls in grades other than Grade 6). Independent of the program, ability leads to perceived time discounted future wage and non-wage benefits, where these non-program benefits are likely to be concave increasing in academic ability, and the cost of exerting study effort is a convex increasing function.

A related argument suggests that teachers in program schools would also exert more effort than teachers in comparison schools. If teachers face a maximization problem similar to that for students, in which they experience benefits (i.e., ego rents, social benefits in the community, or even gifts from parents) from having more scholarship winners in their class, then they should also increase their work effort. Teachers might also simply find extra effort more worthwhile when their students are putting more effort into their studies. Larger non-monetary costs to shirking for teachers in program school communities – including informal social sanctions on the part of parents or the headmaster – might also lead to increased teacher effort, although note that we do not formally model social sanctions above.

It is possible that such social sanctions could differ across communities as a function of local parent support for the program in which case the merit award would generate larger gains where parents are more supportive. The June 2003 structured interviews with teachers provide evidence on how parental support may have contributed to program success. For instance, one teacher mentioned that after the program was introduced, parents began to “ask teachers to work hard so that [their daughters] can win more scholarships.” A teacher in a different school asserted that parents visited the school more frequently to check up on teachers, and to “encourage the pupils to put on more efforts.”

An equilibrium consists of a time path of effort levels by all students and teachers such that each player’s behavior is optimal given the choices of other players. The introduction of the award can potentially lead to greater study effort among those eligible for the award, among those who will be eligible in future years (if they seek to increase their academic ability to boost future chances of winning), and among other students in the class. These patterns may not hold, however, if student effort levels are substitutes, in which case, students with little chance at the award may free-ride on the effort of classmates who are exerting more effort to compete for the prize. The award can also lead to persistent test gains, since a one-time increase in effort raises future ability.

This framework illustrates how even those individuals in program schools who are ineligible for awards (i.e., boys) or who are eligible but unlikely to win awards (i.e., girls with very low initial academic ability) might benefit from the program, through several possible channels. First, greater effort by classmates could improve the classroom learning environment and boost scores directly through a peer effect. Second, these students could directly benefit from increased teacher effort, to the extent teacher effort benefits the entire class and is not targeted only to the girls with a good chance at winning the merit award. Third, to the extent that the student's own effort complements classmates' and teachers' effort in educational production, even children without incentives might optimally exert additional effort themselves, boosting test scores through a multiplier effect. For example, studying becomes more attractive relative to goofing off or daydreaming in class if the teacher is present in the classroom, and one's classmates are also studying hard and learning (Lazear 2001).

There are several other plausible effects that are not explicitly modeled above. If individuals experience utility benefits from their relative ranking in class, then boys ineligible for the merit award might exert additional effort in order to "keep up with" girls in the class who are exerting more effort (although it is worth noting that this relies on a non-standard assumption regarding individual utility, in particular the concavity of utility in relative scores). Finally, if the merit award boosts school attendance among the grade 6 girls (typically 13-16 years old) striving for the award, and if adolescent boys prefer to attend school when more adolescent girls are also present at school, then the program would increase their school participation as well.

In the empirical work that follows, we focus on reduced form estimation, in other words, the impact of the incentive program on test scores. We also estimate program impacts on multiple possible channels linking individual behavior to test scores – in particular, measures of student and teacher effort, as well as other factors (e.g., student attitudes toward school and self-esteem) that are not explicitly modeled above – to better understand the mechanisms underlying the reduced form estimates.

4. Data and Estimation

4.1 The Dataset

The test score data were obtained from the District Education Offices (DEO) in Busia district and Teso district. Test scores were normalized in each district such that scores in the comparison school sample (girls and boys together) are distributed with mean zero and standard deviation one. The complete dataset with both the cohort 1 and cohort 2 students enrolled in school in January 2001 is called the *baseline sample* (Table 1, Panel B). In the main analysis, we focus primarily on students with complete age and gender information, in schools that did not pull out of the program and for which we have mean school grade 6 baseline 2000 test scores and school ethnic composition, and call this the *restricted sample* (Panel C). Note that average test scores are slightly higher in the restricted sample than in the baseline sample, since the students dropped from the sample are typically somewhat below average in terms of academic achievement, as discussed below.

As discussed above, six of the 127 schools invited to participate decided to pull out of the program, leaving 121 schools. Five additional schools (three in Teso district and two in Busia) with incomplete 2000, 2001, or 2002 exam scores, or missing demographic data were also dropped, leaving 116 schools and 7,219 students in the restricted sample (students in program schools account for 50%).

Attendance data are based on four unannounced checks, one conducted in September or October 2001, and one in each of the three terms of the 2002 academic year. Collected by NGO enumerators, these data record as “present” those baseline students actually in school on the day of the unannounced check. Attendance rates are somewhat below 80% for the baseline sample and slightly over 80% for the restricted sample (Table 1, Panels B and C). We use data from these unannounced checks rather than official school attendance registers, since registers are often unreliable in less developed countries.

Household characteristics are similar across program and comparison schools (Table 2): there are no significant differences in parent education, number of siblings, proportion of ethnic Luhyas, or the ownership of a latrine, iron roof, or mosquito net (using data from the 2002 student surveys), indicating that the randomization was largely successful in creating comparable groups.[13] Further evidence is provided by comparing the 2000 (baseline) test score distributions, which are very similar graphically for cohort 1 girls in Busia (Figure 2). Formally, we cannot reject the hypothesis that average baseline test scores are the same across program and comparison schools, as discussed below.

Another estimation concern is the possibility of cheating on the district exam in program schools, but this appears unlikely for a number of reasons. First, district records from external exam invigilators indicate there were no documented instances of cheating in any sample school during either the 2001 or 2002 exams. Several findings reported below also argue against the cheating explanation: test score gains among cohort 1 students in scholarship schools persisted a full year after the exam competition, when there was no longer any direct incentive to cheat, and there were substantial, though smaller, gains among program school boys ineligible for the scholarship, who had no clear benefit from cheating (although cheating by teachers could still potentially explain that pattern). There are also program impacts on several objective measures of student and teacher effort – most importantly, school attendance measured during unannounced enumerator school visits.[14]

4.2 Sample Attrition

There is a large and statistically significant difference in attrition across program and comparison schools in Teso district, but much less so in Busia. Among cohort 1 students, 82% of baseline students in Busia scholarship schools and 77% in comparison schools took the 2001 exam. Thus there is a small, positive but insignificant point estimate of 5% on the difference between the proportion taking the 2001 exam between scholarship and comparison schools (Table 3, Panel A1). Among cohort 2 students there is almost no difference between the scholarship school students and comparison students in Busia district

[13] This comparison in Table 2 relies on the assumption that the household characteristics (i.e., parent education, fertility, ethnicity, and asset ownership) were not directly affected by the scholarship program by the time surveys were collected in mid-2002, which seems reasonable. There is no analogous survey data from 2001.

[14] Jacob and Levitt (2002) develop an empirical methodology for detecting cheating teachers in Chicago primary schools, which relies on identifying classes where test scores rose sharply in a single year (the year of the cheating) and not in other years, and where many students had suspiciously similar answer patterns. Although we cannot examine the second issue, since we only have total test scores on the district exams, the finding of persistent test score gains in the year following the competition argues against cheating as an explanation for our main result.

who took the 2002 exam (Panel A2). There is more overall attrition by 2002 as some students had dropped out of school, transferred to other schools, or decided not to take the district exam. However, among baseline sample cohort 1 students in Teso, 63% of scholarship school students took the 2001 exam, while the rate for comparison school students is 77% (Table 3, Panel A1), and while only three percent of students in Busia district were in schools that pulled out of the program, fully 12% of students in Teso were in schools that left the program (Table 3, Panels B1 and B2).

Among cohort 1 students, the restricted sample includes 77% of both the baseline program school and comparison students in Busia district (Table 3, Panel C1). In Teso, however, only 54% of program students and 66% of comparison students remain in the restricted sample (Panel C1). Thus attrition rates are much higher and less balanced in Teso district than in Busia, and this may in part explain the different estimated program impacts in these two districts.

Differential attrition between program and comparison schools in Teso district is smaller among cohort 2 students than cohort 1. To perhaps understand why, recall that the 2002 district exams for Teso were canceled in the run-up to Kenyan national elections, and the NGO instead administered its own exam – modeled on standard government exams – in Teso in early 2003. Students did not need to pay a fee to take the NGO exam, unlike the government test, and this is likely to account at least in part for the low levels of attrition for cohort 2 in Teso district.

There is some evidence that the scholarship program led academically weaker students in program schools – who ordinarily would not have paid to take district exams – to take the exam, 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 evidence of this. Figure 3 presents non-parametric Fan locally weighted regressions that show attrition to the restricted sample in 2001 as a function of baseline 2000 test scores for cohort 1 girls in Busia; the analogous non-parametric plots for Teso district girls yield broadly similar patterns (not shown). Lower academic achievement in 2000 typically corresponds with higher attrition overall, but this pattern is much more pronounced for comparison school students. In other words, lower-achieving students in scholarship program schools were more likely to take the 2001 exam (i.e., less attrition) than similar students in comparison schools, and this difference across groups is statistically significantly different than zero in the left tail of the baseline 2000 exam distribution (regression not shown).

Confirming these findings, students who did not take the 2001 exam ("attritors") were somewhat lower achieving students on average at baseline in both Busia and Teso districts (Table 4, Panel A). 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 0.17 standard deviations lower at baseline on average than those who did take the 2001 exams (column (i)-(iv)), but the difference is 0.58 standard deviations in Busia comparison schools (column (ii)-(v)), and the difference is statistically significant – further evidence that a greater proportion of low performing students attrited from comparison schools than from program schools. Taken together, this suggests that program impact estimates are likely to be lower bounds on true effects.

In reference to attrition due to schools pulling out of the program, Teso district students whose schools pulled out of the program, and were in program schools, were typically higher achieving students than those in the comparison schools that pulled out, scoring a massive 1.49 standard deviations higher in 2000 on average (Table 4, Panel B). This is perhaps due to individuals in high-performing Teso program schools feeling more "vulnerable" to the program – since they were more likely to win – than similar individuals in comparison schools, in Teso communities where there was mistrust of the NGO, although it is worth noting that this result is based on the small number of schools that pulled out. Note, however, that one girl in Teso who won the ICS scholarship in 2001 refused the scholarship award (see Figure 1).

To summarize, Teso district primary schools had higher rates of sample attrition than Busia schools in 2001, and the gap in attrition across program versus comparison schools was also much greater in Teso district. Students in schools that pulled out of the program in Teso appear to have somewhat better students than those who participated in the program. These patterns all complicate causal inference in Teso district. In what follows, we focus the main analysis and interpretation on Busia district, where there was no evidence of differential attrition and where few schools pulled out of the program, although we also present the main results for the full sample of both districts.

4.3 Estimation Strategy

The main estimation equation is:

(1) $$TEST_{ist} = Z_{ist}{}'\beta + (Z_{ist} * T_s)'\gamma + X_{ist}{}'\delta + \mu_s + \varepsilon_{ist}$$

$TEST_{ist}$ is the test score for student i in school s in year t. Z_{ist} is a vector of indicator variables for each cohort and year (i.e., cohort 1 in year 1, cohort 1 in year 2, etc.), and T_s is an indicator for the program schools. In specifications where the goal is to estimate the overall program impact across all cohorts and years, we exclude the $Z_{ist}*T_s$ term and instead estimate the coefficient estimate on the treatment indicator. X_{ist} is a vector of other explanatory variables, including student age, the mean school grade 6 baseline test score, and controls for school ethnic composition. Error terms are assumed to be independent across schools, but are allowed to be correlated across observations within the same school. The disturbance terms consist of μ_s, a school effect perhaps capturing common local or headmaster characteristics, and an idiosyncratic term, ε_{ist}, which may capture unobserved student ability or shocks. The non-parametric locally weighted regression technique in Fan (1992) allows us to estimate average program impacts across individuals with different baseline scores. We use similar methods to estimate impacts on behavioral channels (e.g., school attendance) potentially linking the program to test scores.

5. Empirical Results

5.1 Academic Test Score Impacts

The scholarship program raised test scores by 0.12 standard deviations (standard error 0.05) overall among boys and girls in 2001 and 2002, pooling Busia and Teso districts and students of both cohorts (Table 5, Panel A, regression 1), and this gain is statistically significant at 95% confidence. The estimated impact of the program is larger for girls, as expected, with a sizeable average gain of 0.17 standard deviations (standard error 0.06, statistically significant at 99% confidence, regression 2) overall in both Busia and Teso, while the average effect for all boys is 0.09 (not statistically significant). Boys score much higher than girls on average, with a gender gap of 0.34 (standard error 0.04).

The estimated overall effect, without any attrition corrections, for girls and boys together, is considerably larger for Busia district (0.20, standard error 0.07, Table 5, Panel B, regression 1) than for Teso (-0.02, standard error 0.07, regression 2). This is consistent both with the hypothesis that winning a scholarship was less desirable in Teso due to mistrust of the NGO or lack of social prestige associated with winning the award, and with the possible bias due to sample attrition in Teso.

To address non-random sample attrition, we construct non-parametric bounds on the overall program effect in Teso using the trimming method in Lee (2002); it does not make sense to construct bounds for Busia schools since there was no differential attrition across program groups there. The treatment effect bounds for cohort 1 girls in Teso district are wide, ranging from -0.24 standard deviations as a lower bound up to 0.22 standard deviations as an upper bound. Thus while we cannot rule out that the program in fact had a positive impact in Teso, the high and unbalanced attrition in Teso district makes it difficult to draw firm conclusions about the effect of the program there.

Although it is difficult to convincingly address attrition bias, a simple imputation exercise, where missing values for the 2001 test score are filled with a predicted score, as a function of the 2000 baseline exam score, suggests that the expected program impact for cohort 1 girls in Teso district in the absence of attrition would be slightly more positive, but still small and not statistically significantly different than zero (-0.02, standard error 0.08 – regression not shown). As further evidence that effects were small in Teso, the estimated program impact among cohort 2 Teso girls in 2002 – a subsample for which there was little differential attrition across treatment groups (Table 3, Panel C2) – is near zero and statistically insignificant (estimate 0.00, standard error 0.11 – regression not shown).

Whatever interpretation is given to the Teso district results – either no actual program impact, or simply unreliable estimates due to attrition – the fact remains that the program was less successful in Teso, at a minimum in the sense that fewer schools chose to take it up. It remains unclear whether the problems encountered in Teso district would have arisen in the absence of the lightening tragedy of 2001, and whether they would arise in other settings.[15]

We next separately estimate effects for girls and boys across cohorts and years. The program effect for girls competing for the scholarship is 0.29 standard deviations in the restricted sample of girls in cohort 1 in Busia (competing in 2001), and 0.21 for cohort 2 in 2002 (Table 5, Panel C, regression 1), and in both cases the effects are significantly different than zero at 95% confidence. The main result for cohort 1 girls in Busia is robust to using the change in test scores between 2000 and 2001 as the dependent variable (coefficient estimate 0.20, standard error 0.12 – regression not shown). These are large impacts: to illustrate with previous findings from Kenya, the average test score for grade 7 students who take a grade 6 exam is approximately one standard deviation higher than the average score for grade 6 students (Glewwe et al 1997), and thus the estimated program gains roughly correspond to an additional 0.21-0.29 grades of primary school learning.

Other explanatory variables have expected effects. The baseline mean school grade 6 test score in 2000 is significantly positively associated with the 2001 test score (Table 5, Panel C)[16]. Being one year older decreases girls' test scores by 0.09 standard deviations; in Kenya, older students within the same grade have usually either repeated a grade or entered school later than others. The ethnic composition controls have some predictive power, and a higher proportion of ethnically Teso students in Busia district schools is associated with higher test scores. This is consistent with the hypothesis that higher quality primary schools attract more ethnically diverse student populations, as argued in Miguel (2001). Program impact estimates are similar if these explanatory variables are excluded, although estimates are less

[15] To potentially disentangle the effect of being in a Teso district school from the effect of the lightning strike (in a specification that pools the Busia and Teso data for all girls and boys), we included an indicator variable for Teso district, and an interaction of the Teso indicator with the program indicator, as well as an indicator for schools located with 6 km of the lightning strike, and the interaction of this distance term with the program indicator. The coefficient estimate on the lightning distance and program indicator interaction term is negative but not statistically significant (-0.05, standard error 0.09 – regression not shown), while the coefficient estimate on the Teso-program interaction term remains negative and marginally significant (-0.25, s.e. 0.11). Still, these results do not rule out that program impacts in Teso district might have been positive in the absence of the lightning strike.

[16] Including *individual* baseline test scores from 2000, for Busia girls in cohort 1 (for whom this data is available), does not substantially change the results although it does reduce estimates somewhat: the estimated program impact in year 1 becomes 0.19 (standard error 0.12), and for year 2 the post-program impact is 0.22 (standard error 0.08).

precise.[17] Estimates are largely unchanged when individual demographic controls collected in the 2002 student survey – including parent education and household asset ownership – are included as explanatory variables.[18] Interactions of the program indicator with these household socioeconomic proxies, including parent education levels, are not statistically significant at traditional confidence levels (regressions not shown), suggesting that test scores did not increase by a larger amount for students from higher socioeconomic status households (note that although the program had a similar test score impact across socioeconomic backgrounds, students with more educated parents nonetheless were disproportionately likely to win because they had higher baseline scores). Similarly, neither the average baseline school test score, nor the proportion of female teachers in the school significantly affects average program impacts (regressions not shown).

The scholarship program not only raised test scores when it was first introduced in 2001, but also continued to boost scores of cohort 1 girls during 2002: the point estimate in year two is 0.28 (standard error 0.08, Table 5, Panel C regression 1) for the restricted sample, providing additional evidence that the program had lasting effects on learning, rather than simply being due to cheating or cramming for the 2001 exam. There is further evidence on longer-term impacts in Busia from the ICS exam, administered in February 2003. Although originally conducted in order to obtain test scores in Teso district that could be used to determine program winners (after the 2002 Teso district exams were canceled), they were also conducted in the Busia district sample schools. In the standard specification (like those in Table 5), the impact of the program on Busia cohort 1 girls in 2003 was 0.19 standard deviations (standard error 0.07, statistically significant at 99% confidence), and the gain for Busia cohort 2 girls is also statistically significant at 0.15 standard deviations (standard error 0.08 – regressions not shown). Though average

[17] For instance, the program impact for Busia cohort 1 girls is 0.27 standard deviations in this case and the standard error rises to 0.19, while the program impact for cohort 2 rises to 0.28 with standard error 0.17 – regressions not shown. In contrast, standard errors fall considerably when disturbance terms are not clustered at the school level; for instance, the standard error on the overall effect for girls and boys in Busia and Teso (as in Table 5, Panel A) decreases from 0.05 to 0.02 (regression not shown). Estimated impacts for cohort 1 Busia girls are similar when school average values are used, rather than individual micro-data (coefficient estimate 0.23, s.e. 0.09 – not shown).

[18] These are not included in the main specifications because they were only collected for a subsample of students, those present on the day of 2002 survey administration, and this would thus reduce the sample size and change the composition of students somewhat.

program impacts fall somewhat for cohort 1 in the second year after the competition – from 0.29 in the year of the competition, to 0.28 in the year following the competition, to 0.19 at the start of the second year after the competition – program impacts remain remarkably persistent, and there are no statistically significant differences across years[19].

We next focus on graphical representations of test score impacts for Busia girls. Baseline scores are nearly identical across scholarship and comparison schools (Figure 2). The vertical line indicates the minimum score that won the scholarship in 2001. The score distribution shifts to the right in program schools for cohort 1 in year 1 (Figure 4), cohort 1 in year 2 (Figure 5), and cohort 2 in year 2 (Figure 6). [20] The largest gains appear to be near or right below the minimum winning score threshold, consistent with the view that the students exerting the most additional effort were those who believed that additional effort could make the greatest difference in their chances to win the award.

These figures above do not allow us to determine the magnitude and statistical significance of program effects at different regions of the initial test score distribution, but Figure 7 presents a non-parametric Fan locally weighted regression that shows the scholarship program impact for Busia cohort 1 girls as a function of their individual test score in 2000. Girls just below the winning threshold had large test score gains. High-achieving girls in 2000 had the smallest increases in 2001 test scores, perhaps since girls with the highest baseline scores could exert less effort and still remain above the threshold to win the scholarship, or perhaps in part because the highest achieving girls at baseline were already exerting something close to their "maximum" effort. There are also marked gains at the bottom of the baseline test score distribution for girls, suggestive evidence of positive spillover benefits of the program even among girls with little realistic chance of winning. However, it is impossible to reject the hypothesis that gains at the bottom of the baseline distribution are the same as gains elsewhere due to limited statistical power.[21]

[19] We cannot reject the hypothesis that program effects in year 1 are equal to either the 2002 or 2003 post-program effects for cohort 1 girls in Busia (p-values of 0.96 and 0.38, respectively).
[20] These figures use a quartic kernel and a bandwidth of 0.7.
[21] The program had somewhat larger effects on scores in mathematics, science, and geography/history than in English and Swahili (Appendix Table B), but overall differences by subject are not statistically significant.

Field interviews conducted in July 2002 indicate that students actively competed for the scholarship when it was offered. One headmaster reported that the program "awakened our girls and was one step towards making the girls really enjoy school."[22] 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 plausible that this spirit of competition drove some girls to work harder, providing utility benefits beyond the direct program monetary rewards.

Boys in Busia district program schools also have higher test scores than those in comparison schools despite not being eligible for the scholarship themselves. The overall effect for Busia boys in both cohorts 1 and 2 was 0.16 (standard error 0.08, regression not shown), which is nearly significantly different than zero at 95% confidence. Cohort 1 gains in 2001 were even larger, at 0.21 standard deviations (standard error 0.09) in the restricted sample (Table 5, Panel C, regression 2). Since boys and girls share the same classroom, boys are likely to benefit if teachers exert more effort as a result of the program. In the first year, it is also possible that some boys were confused as to whether they too were eligible for the scholarship, although the survey data presented above (section 2.2) suggests that if it exists, this effect is likely to be very small. In the second year of the program, there are again positive, though not statistically significant, program impacts for boys – although we cannot reject that effects for boys are the same in 2001 and 2002. Overall, there is no evidence the girls' scholarship program discouraged or demoralized boys, at least in terms of their academic performance.

The 2002 field interviews suggest that a desire to compete with girls drove some boys to study harder. To the extent that this "gendered" competition was an important determinant of boys' gains in program schools, it is an open question how large externality gains would be under an alternative program that targeted boys rather than girls, or in which they competed against each other for the same awards. Gneezy et al (2003) provide experimental laboratory data that females and males may sometimes react differently to competition, with females performing better when competing against other females than when competing against males, while males perform equally well regardless of the competition's gender.

[22] Source: authors' field notes, July 15, 2002.

The focus so far has been on the first moment of the test score distribution. The data suggest that there was a small overall increase in test score variance in program schools relative to comparison schools, but differences are minimal and generally not statistically significant (Table 6).[23] The changes in variance over time in program versus comparison schools are similarly minimal for boys (not shown).

5.2 Channels: School Participation, Behaviors and Attitudes

It is useful to explore potential channels for test score gains, since some mechanisms, such as cheating or increased coaching, might raise test scores without improving underlying learning. Using the same set of educational outcomes measures as Glewwe et al. (2003), we find starkly different patterns. School attendance and medium-run test score effects are two indicators of effort aiming to increase long-run human capital, while we treat extra test preparation sessions as indicators of effort to increase test scores. The scholarship program significantly increased student attendance (measured during unannounced enumerator visits) in 2001 and 2002 in Busia district (Table 7, Panel A, regression 1): pooling cohorts 1 and 2 in the Busia restricted sample and measuring the effect of the scholarship on overall attendance yields a coefficient estimate of 5 percentage points (standard error 2 percentage points), which is statistically significant at 95% confidence, and corresponds to approximately a 30% reduction in student absenteeism. These attendance gains indicate that program school students are exerting extra effort in one important and easily measured dimension. There are statistically insignificant effects in Teso district (regression 2), and the estimated student school attendance gain pooling Busia district and Teso district together for girls and boys is insignificant at traditional confidence levels (2 percentage points, standard error 2 percentage points – regression not shown).

The program increased the average likelihood of school attendance by 6 percentage points among girls in cohort 1 in 2001, and by 10 percentage points among cohort 2 in 2001 (a pre-program effect), and estimated gains in 2002 are also positive but smaller (Table 7, Panel B, regression 1). The pattern of

[23] The slight (though insignificant) increase in test score inequality in program schools is inconsistent with one particular naïve model of cheating, in which program school teachers simply pass out test answers to their students. This would likely reduce inequality in program relative to comparison schools. We thank Joel Sobel for this point.

attendance gains for Cohort 1 Busia girls in 2001 is similar to the pattern in test score gains for that subsample, with large gains near the winning test score threshold and also at the bottom of the test score distribution (Figure 8). Busia district boys in scholarship schools show similar effects, with larger gains in 2001 than in 2002 (Table 7, Panel B, regression 2). It is unclear exactly why attendance gains are larger in 2001 than in 2002 but perhaps there was greater enthusiasm at the start of the program – especially if some students mistakenly over-estimated the effect that extra study effort would have on their chances of winning the award in year 1, and this estimate were revised downward by cohort 2 in year 2.[24] The attendance gains among cohort 2 in 2001 might be due to anticipation of the future competition (they competed in 2002). Program school attendance impacts were not significantly different between school terms 1, 2 and 3 in 2002 (not shown), so there is no evidence that gains were largest in the period immediately preceding exams, due to cramming, for instance.

Teachers in Busia program schools were five percentage points more likely to be present at school than comparison school teachers during 2002, reducing overall teacher absenteeism by approximately one-third (Table 7, Panel C).[25] Estimated program impacts are not statistically significantly different as a function of teacher gender or experience (regressions not shown).

The 2002 student survey collected information on educational inputs, study habits, and attitudes that may have affected school performance, to partially capture other dimensions of study effort. As the survey was administered in mid-2002, cohort 1 girls had already competed for the scholarship when they filled out the questionnaire, while cohort 2 girls had not yet competed for it. 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 (although note there is no effect for cohort 2). However,

[24] There is no significant program effect on the likelihood of dropping out of school by 2002, although the point estimate goes in the expected direction (regression not shown).

[25] These results are for all teachers in the schools. It is difficult to distinguish between teacher attendance in grade 6 versus other grades, since the same teacher often gives a class (i.e., mathematics) in several different grades in a given year, and the data were recorded on a teacher by teacher basis rather than by grade and subject, unfortunately. Thus, it remains possible that average teacher attendance gains would be even larger for grade 6 classes alone.

there is no impact on the likelihood that program school students sought out extra school coaching ("preps" in Kenya), handed in homework, were called on by the teacher in class, or did fewer chores at home (Panel A). In the case of chores, the estimated zero impact suggests there was minimal cost of the program in terms of lost production at home, suggesting any increased study effort may have come out of children's leisure time.

In terms of educational inputs such as the number of new textbooks or exercise books available at home, there are no significant gains in program schools (Table 8, Panel B), although all six point estimates presented in the panel are positive, providing suggestive evidence in favor of some increased parental investments in child school supplies. Further evidence is provided by a specification that pools all cohort 1 and 2 girls and boys in Busia district, and finds an increase of 0.22 additional new exercise books or textbooks at home in program schools (standard error 0.19, regression not shown)[26].

There is no convincing evidence for any positive or negative impacts on attitudes toward education, for instance, thinking of oneself as a "good student", or preferring school activities to non-school activities, based on survey responses (Table 8, Panel C). This is evidence against the view that external incentives dampened intrinsic motivation in this context.

To summarize, overall there is evidence of increased student and teacher effort (reflected in school attendance and textbook use at home), suggestive evidence of increased parental educational investments, but no evidence of the adverse attitude changes emphasized by some psychologists.

5.3 Regression Discontinuity Estimates of the Impact of Winning the Scholarship

The impact of winning the program is estimated using a regression discontinuity method, which compares the 2002 outcomes of girls who barely won the scholarship in 2001 (their 2001 test score was slightly above the winning threshold) to girls who barely lost. In practice, 2001 test score polynomials (linear, quadratic, and cubic terms, and these terms interacted with the program school indicator) are included to control for any smooth underlying relationship between the 2001 test score and later outcomes, and an

[26] There was no program impact on classroom inputs, including desks and flipcharts, which we estimate using data collected during 2002 classroom observations (regressions not shown).

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 estimate both the impact on scholarship winners and any possible “demoralization” experienced by non-winners.

The coefficient estimate on the interaction term between the program school indicator variable and the indicator for scoring above the winning scholarship threshold captures the impact of winning the scholarship (coefficient *f* in equation 2):

$$(2) \qquad Y_{is} = a + X_{is}'b + c \cdot WIN_{is}^{2001} + d \cdot T_s + f \cdot (WIN_{is}^{2001} * T_s)$$
$$+ \sum_{z=1}^{3} \left\{ g_z \left(TEST_{is}^{2001}\right)^z + h_z \left(TEST_{is}^{2001} * T_s\right)^z \right\} + u_s + e_{is}$$

There are large and statistically significant school participation gains in 2002 among scholarship winners (point estimate 0.20, statistically significant at 99% confidence – Table 9, regression 1), evidence that paying girls’ school fees has a large positive impact on school participation. The school participation gains for winners in program schools are zero in term 1 (0.00, standard error 0.05), but they grow much larger in terms 2 and 3 (at 0.35, standard error 0.10 and 0.27, standard error 0.10, respectively), suggesting widening school participation impacts for winners versus non-winners through time. The overall point estimate is somewhat smaller, but remains large, positive and statistically significant when the effect is estimated only using program school students (0.09, standard error 0.03 – regression not shown). The leading explanation is that paying winners’ school fees led to higher school participation in 2002. Note, however, that there are no significant 2002 school participation gains for winners in Teso district schools (not shown), and it is unclear what explains this difference across the two districts.

However, winners are no more likely than non-winners in Busia district to claim that they think of themselves as “good students” (Table 9, regression 2), nor is there a significant impact on preferences for school relative to other activities (regression 3), on the number of textbooks or exercise books owned (regressions not shown), or on 2002 test score performance (regression 4), and in this last case the point estimate for winners is negative though not statistically significant. There is also no evidence of negative demoralization effects for non-winners in program schools in terms of any of the same four outcomes (this effect is captured by the program school indicator in Table 9).

6. Program Cost-effectiveness

We compare the cost-effectiveness of five programs that have recently been conducted in the study area – the girls' merit scholarship program that is the focus of this paper, the teacher incentive program discussed above (Glewwe et al. 2003), a textbook provision program (Glewwe et al. 1997), flip chart program (Glewwe et al. 2004), and deworming program (Miguel and Kremer 2004) – and conclude that providing incentives for students are a particularly cost-effective way to boost test scores (Table 10).

The average test score gain in girls' merit scholarship program schools, for both female and male students in Busia and Teso districts in both years of the program, is 0.12 standard deviations (Table 5), while the comparable gains for teacher incentive program schools over two years was smaller, at 0.07 standard deviations, and for textbook program schools the average gain was only 0.04 standard deviations. The test gains in the teacher incentive program were concentrated in the year of the competition, and they fell in subsequent years. The program which provided medical deworming treatment to Busia district primary schools did not produce statistically significant test score impacts, nor did the program which provided flip charts to primary schools. Since the cost per test score gain in these two programs is infinite given our zero estimated impacts, we do not focus on them below.

One issue in a cost-effectiveness analysis is whether to treat all payments under the program as social costs or whether to consider some of them as transfers. In column 4 of Table 10, we report "education budget cost effectiveness" which shows the test score gain per pupil, divided by program costs per pupil. This is the relevant calculation for an education policymaker, in other words, maximizing test gains with a given budget. From the standpoint of a social planner, however, some of the payments to families in the scholarship program, and to teachers in the teacher incentive program, could be considered as transfers. If these are seen as pure transfers, the social cost is simply the deadweight loss involved in raising the necessary funds. In calculating "social cost effectiveness" we follow a rule of thumb often used in wealthy countries and treat the marginal cost of raising one dollar as 1.4 dollars (Ballard et al. 1985). In order to make the education budget and social cost effectiveness figures comparable, we also multiply all costs in the education budget calculations by 1.4 to reflect likely tax distortions.

It is worth noting that the transfer to families in the merit scholarship program, and to teachers in the teacher incentive program, is the net benefit to them after making allowances for any disutility of their increased effort. Assuming that students and teachers are rational, the total additional effort exerted by participants should be less than or equal to the value of the rewards. Thus, the education budget cost effectiveness calculation yields an upper bound on the true social cost of these two programs (Table 10, column 4), while a lower bound is generated by treating the entire payment as a transfer (column 5).

Using project cost data from NGO records, in terms of the social cost effectiveness calculation, the per pupil cost per 0.1 standard deviation gain in the average academic test score is US$1.41 for the girls scholarship program, and nearly identical at US$1.36 per 0.1 standard deviation average gain for the teacher incentive program, but costs are much higher for the textbook program, at $5.61 per 0.1 s.d. gain (Table 5, column 5). Recall that the teacher incentive program did not produce lasting test score impacts, and there is evidence of “teaching to the test” rather than effort directed at human capital acquisition in the teacher incentive schools; as a result, the long-term impacts of the girls scholarship program are likely to be much larger than the teacher incentive program. If attention is restricted to Busia district, where the girls’ scholarship program was well-received by residents, the social cost per 0.1 s.d. gain per pupil falls to US$0.71, making student merit awards much more cost-effective way to boost student scores than the other programs. Merit awards are also more cost effective in raising test scores than textbook provision, deworming, or flip charts under the education budget calculation (column 4).

The estimates for both the girls scholarship program and the teacher incentive program do not include costs associated with administering academic exams in the schools (exam scores are an integral part of these programs, since they provide the information necessary for awarding prizes). Including testing costs, the social cost per s.d. average test score gain nearly doubles for the girls scholarship schools, from US$1.41 to US$2.78, and more than doubles from US$1.36 to US$3.70 for the teacher incentive program. In both cases these programs remain far more cost-effective than the textbook program. Once again restricting attention to Busia district alone, the per pupil social cost per one standard deviation average test score gain is only US$1.53. Many countries, like Kenya during the study period, already carry out regular standardized testing of primary school students, and in which case the additional exam costs are not necessary, and the previous estimates are the relevant ones.[27]

An additional factor in favor of merit scholarships as opposed to teacher incentives is that the distributional impact is likely to be much more favorable, as the scholarship program provides rewards to pupils instead of teachers, who tend to be well-off in rural Kenya. The scholarship program also generated large impacts on pupil school attendance. Attendance may be considered a benefit in addition to the test score gains, and these benefits are not considered in the above cost calculations . Similarly, it is also likely that scholarship winners have high returns to additional education, and to the extent that winners obtain more education than they would have otherwise, this yields additional program benefits.

7. Conclusion

Merit-based scholarships have historically been an important part of the educational finance system in many countries. Yet many educators have opposed merit scholarships on equity grounds, fearing that they would disproportionately flow to pupils from better-off families (Orfield 2002). The evidence we present suggests that such incentive programs can raise test scores not only for recipients but also for others. A merit-based scholarship for Kenyan adolescent girls had a large positive effect on test scores in both years of the scholarship competition. Initially low-achieving girl students, and boys (who were ineligible for the award), both show considerable test score gains, providing evidence for positive classroom spillovers.

---

[27] Several other studies from less developed countries present analogous cost-effectiveness information, allowing us to compare the girls scholarship program to other schooling interventions. One notable example is a recent study by Banerjee et al. (2003), who find that a remedial education program in India that provided extra teachers to primary school classes raised test scores by approximately 0.2 standard deviations, at a cost of US$2 per child per year; incorporating the deadweight loss and transfer adjustments, this yields a cost of roughly US$0.40 per 0.1 standard deviation gain in average test scores. Unfortunately, these estimates do not include the costs of testing, which is necessary to identify students who require remedial education, and it is unclear how much this would add to costs. Moreover, the remedial education program in Banerjee et al. (2003) relied on hiring local high school graduates at a salary of US$10-15, per month. While Banerjee et al. report that there are many recent high school graduates, usually female, willing to take on temporary jobs at this salary in India, it is not clear that many people with these qualifications would be available in rural Kenya.

Though there are large, significant and robust average program 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. Our inability to find these effects may be due to differential sample attrition that complicates the econometric analysis in Teso schools, or it may reflect the lower value placed on the merit award there – especially in the aftermath of the tragic 2001 lightning strike. The different program impacts in Busia and Teso suggest community support is important.[28]

One way to target more of the benefits of merit scholarships to the poor would be to restrict the scholarship competition to needy schools, regions, or populations, or alternatively to conduct multiple competitions, each restricted to a small geographic area. For instance, if each Kenyan district (or division, an even small administrative unit) awarded merit scholarships to its residents independently of other districts, children would only compete against others who live in the same area, where many households live in broadly comparable socioeconomic conditions. This would allow children even in poor rural areas to have a reasonable shot at an award. To the extent that such a policy would put more students near the margin of winning a scholarship, it would presumably have greater incentive effects.

With the introduction of universal free primary education in 2003, secondary school scholarships have become more prominent on the Kenyan policy agenda. Universal free secondary education is infeasible in Kenya due to the high costs of secondary schooling, estimated at US$580 per pupil per year (World Bank 2004), which if funded for all children of secondary school age would be 18% of Kenyan GDP. The results of this paper suggest that introducing merit scholarships for secondary school in disadvantaged areas, or among disadvantaged groups (such as orphans in Kenya), might lead to improvements in average academic performance, as well as providing access to further education for some bright children of modest means.

---

[28] The magnitude and persistence of girls scholarship program effects on educational, labor market, and fertility outcomes in the long-run remains an open question. We have collected detailed contact information for sample individuals, and in the future plan to follow-up and survey these individuals as they enter adulthood. These findings will ideally allow us to estimate long-run payoffs of increased primary school learning for rural Kenyans.

References

Acemoglu, Daron, and Joshua Angrist. (2000). “How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws”, *NBER Macroeconomics Annual,* 9-59.

Akerlof, George, and Rachel Kranton. (2003). “Identity and Schooling: Some Lessons for the Economics of Education,” *Journal of Economic Literature,* 40, 1167-1201.

Angrist, J. and V. Lavy (2002). "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.

Ballard, Charles L., John B. Shoven, and John Whalley. (1985). "General Equilibrium Computations of the Marginal Welfare Cost of Taxes in the United States," *American Economic Review*, 75(1), 128-138.

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

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.

Central Bureau of Statistics. (1999). *Kenya Demographic and Health Survey 1998*, Republic of Kenya, Nairobi, Kenya.

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

Cooper, Russell, and Andrew John. (1988). "Coordinating Coordination Failures in Keynesian Models", *Quarterly Journal of Economics*, 103(3), 441-463.

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*.

Cornwell, Christopher M., Kyung Hee Lee, and David B. Mustard. (2003). "The Effects of Merit-Based Financial Aid on Course Enrollment, Withdrawal and Completion in College", unpublished working paper.

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.

Duflo, E. (2001). "The Medium Run Effects of Educational Expansion: Evidence from a Large School Construction Program in Indonesia." Unpublished manuscript.

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

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

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. (1997). "Textbooks and Test scores: Evidence from a Prospective Evaluation in Kenya", unpublished working paper.

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

Glewwe, Paul, Michael Kremer, Sylvie Moulin, and Eric Zitzewitz. (2004). "Retrospective v. Prospective Analysis of School Inputs: The Case of Flip Charts in Kenya." forthcoming, *Journal of Development Economics*.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. (2003). "Performance in Competitive Environments: Gender Differences", *Quarterly Journal of Economics*, 118, 1049-1074.

Government of Kenya, Ministry of Planning and National Development. (1986). *Kenya Socio-cultural Profiles: Busia District*, (ed.) Gideon Were. Nairobi.

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

Hanushek, Erik. (2003). "The Failure of Input-based Schooling Policies", *Economic Journal*, 113, 64-98.

Jacob, Brian, and Steven Levitt. (2002). "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating", *NBER Working Paper #9413.*

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.

Lazear, Edward P. (2001). "Educational Production", *Quarterly Journal of Economics*, 116(3), 777-804.

Lee, D. S. (2002). "Trimming the Bounds on Treatment Effects with Missing Outcomes." *NBER Working Paper #T277*.

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.

Leuven, Edwin, Hessel Oosterbeek, Bas van der Klaauw. (2003). "The Effect of Financial Rewards on Students' Achievement: Evidence from a Randomized Experiment", unpublished working paper, University of Amsterdam.

Lucas, Robert E. (1988). "On the Mechanics of Economic Development", *Journal of Monetary Economics*, 22, 3-42.

Miguel, Edward. (2001). "Ethnic Diversity and School Funding in Kenya", U.C. Berkeley CIDER Working Paper #C01-119.

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

Moretti, Enrico. (2004). "Workers' Education, Spillovers and Productivity: Evidence from Plant-level Production Functions", *American Economic Review*, 94(3).

Nurmi, J. (1991). "How Do Adolescents See Their Future? A review of the development of future orientation and planning", *Developmental Review*, 11:1-59.

Orfield, Gary. (2002). "Foreward", in Donald E. Heller and Patricia Marin (eds.), *Who Should We Help? The Negative Social Consequences of Merit Aid Scholarships* (Papers presented at the conference "State Merit Aid Programs: College Access and Equity" at Harvard University). Available online at: http://www.civilrightsproject.harvard.edu/research/meritaid/merit_aid02.php.

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

Summers, Anita A., and Barbara L. Wolfe. (1977). "Do Schools Make a Difference?" *American Economic Review*, 67(4), 639-652

World Bank. (2002). *World Development Indicators* (www.worldbank.org/data).

World Bank. (2004). *Strengthening the Foundation of Education and Training in Kenya: Opportunities and Challenges in Primary and General Secondary Education*. Nairobi.

Figure 1: Map of Busia District and Teso District, Kenya, with location of Girls Scholarship Program Schools

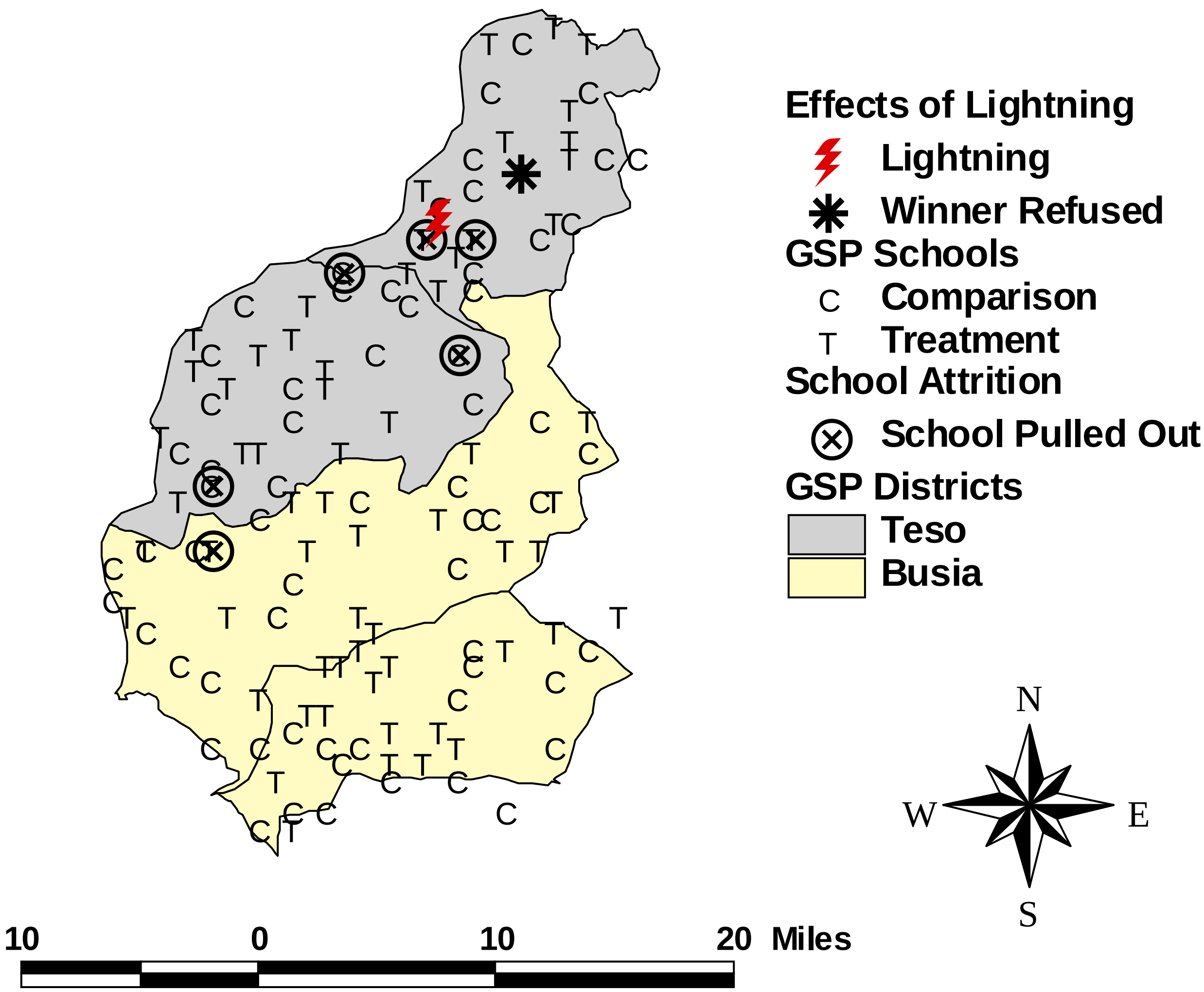

Figure 2: Baseline (2000) Test Scores, Busia Girls Cohort 1
(Non-parametric kernel densities)

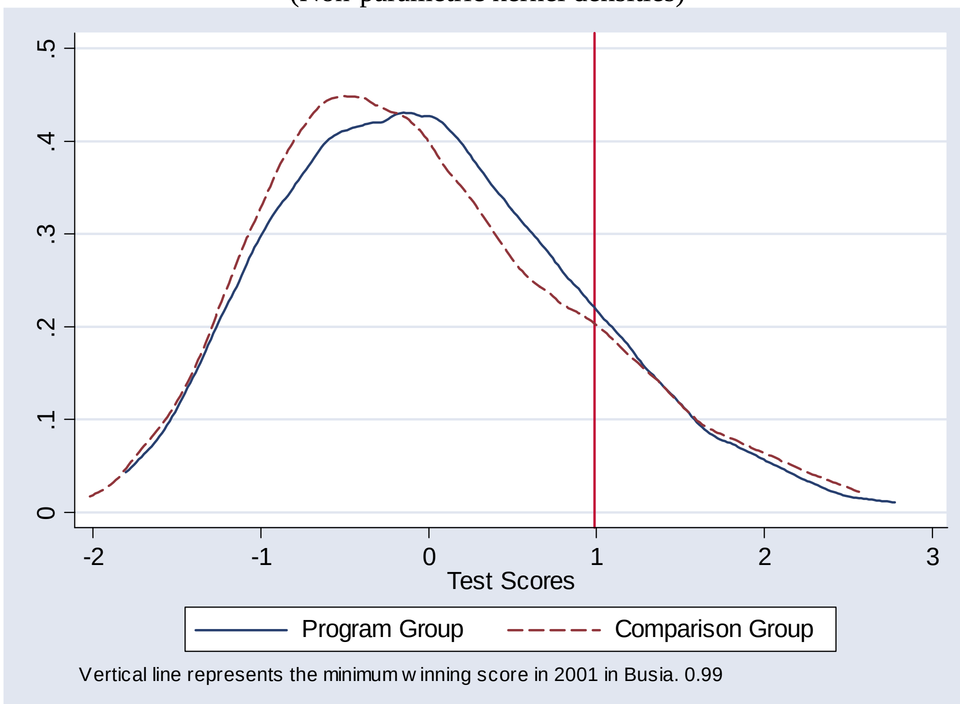

Figure 3: Attrition to the 2001 Restricted Sample by Baseline (2000) Test Score, Busia Girls Cohort 1
(Non-parametric Fan locally weighted regressions)

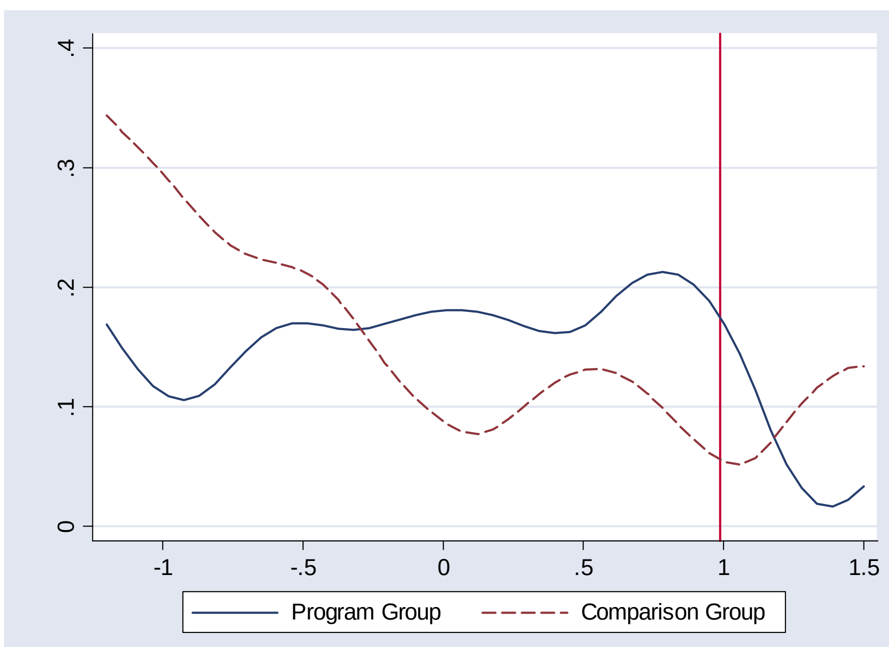

Figure 4: Year 1 (2001) Test Score Impacts, Busia Girls Cohort 1

(Non-parametric kernel densities)

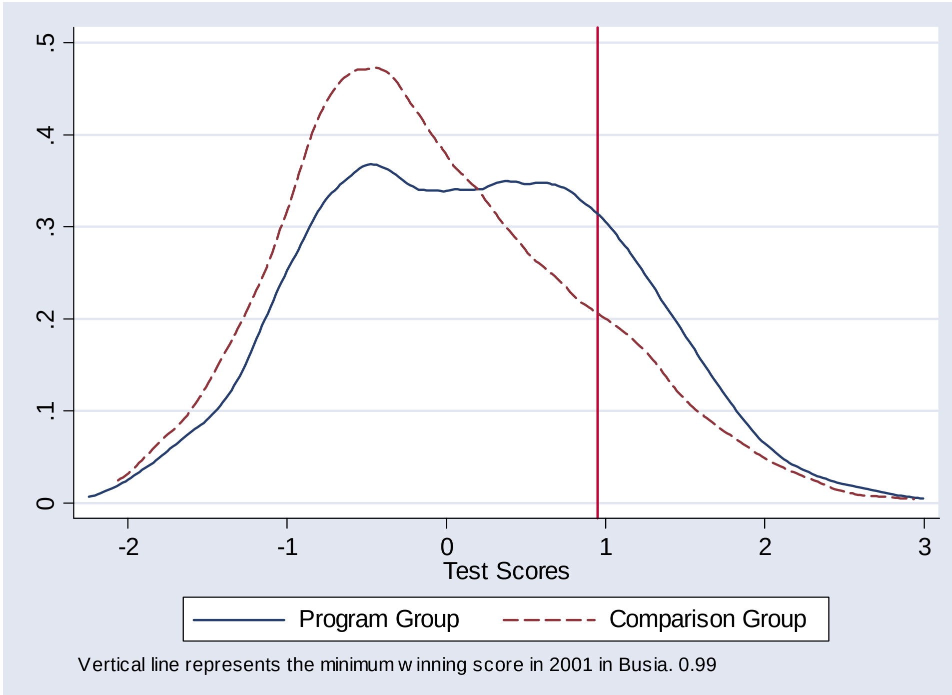

Figure 5: Year 2 (2002) Test Score Impacts, Busia Girls Cohort 1
(Non-parametric kernel densities)

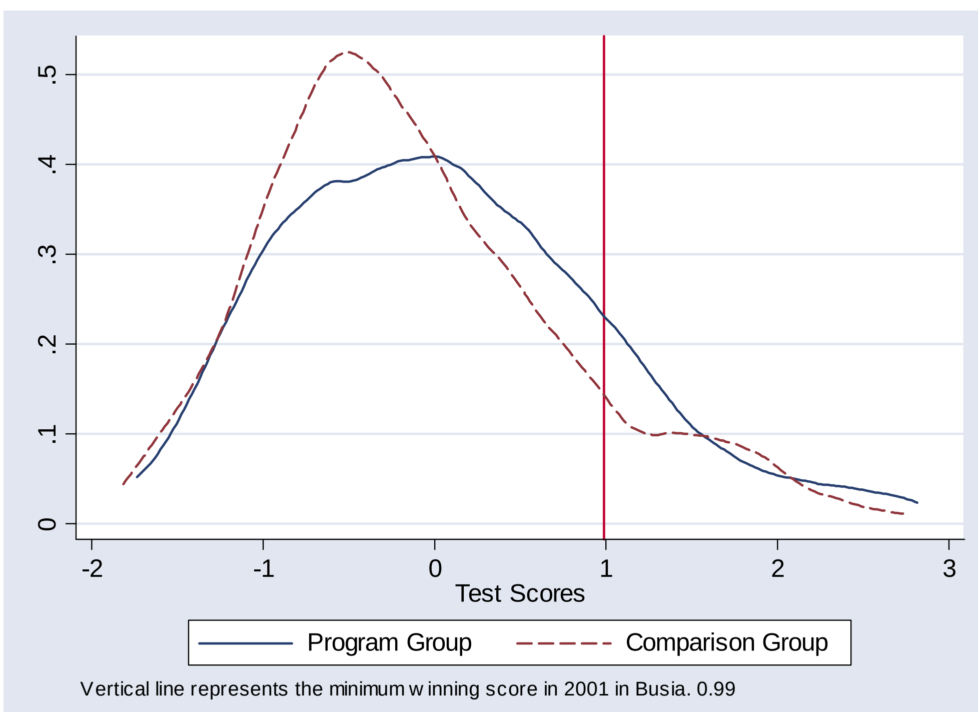

Figure 6: Year 2 (2002) Test Score Impacts, Busia Girls Cohort 2

(Non-parametric kernel densities)

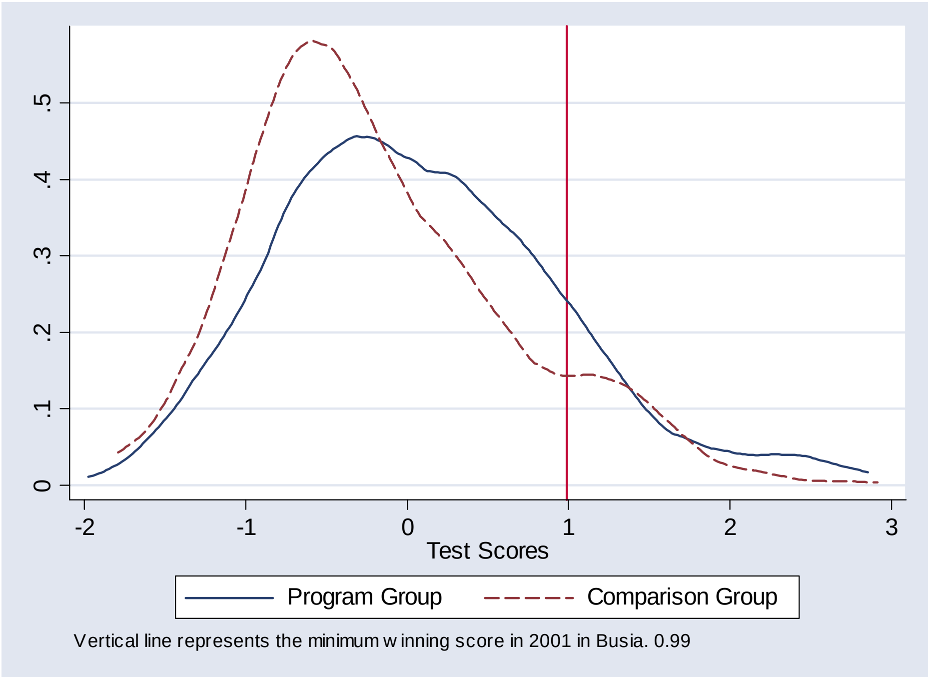

Figure 7: Year 1 (2001) Test Score Impacts by Baseline Test Score, Busia Girls Cohort 1: Difference between Program Schools and Comparison Schools
(Non-parametric Fan locally weighted regression)

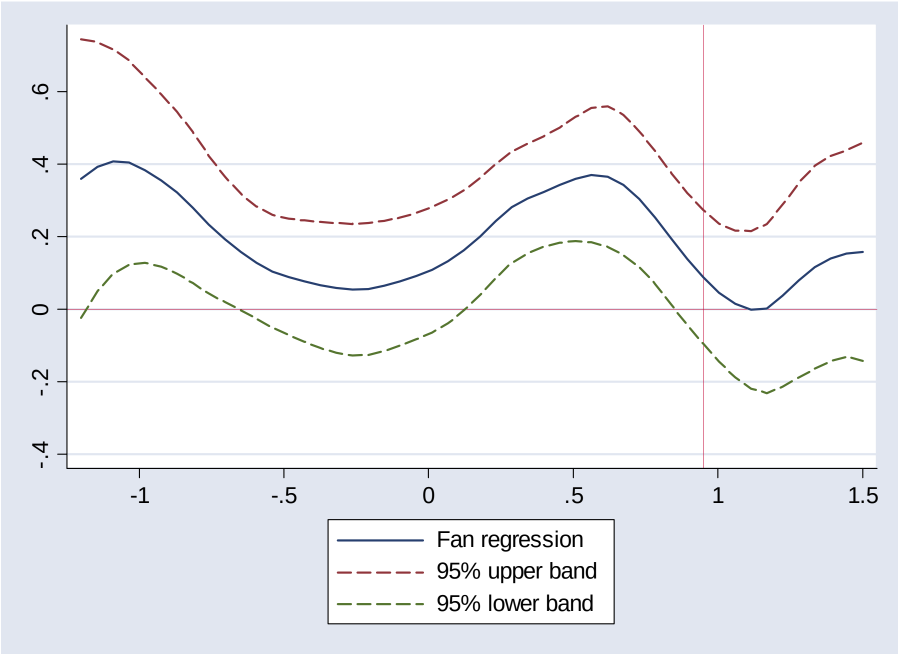

Figure 8: Year 1 (2001) School Participation Impacts by Baseline Test Score, Busia Girls Cohort 1: Difference between Program Schools and Comparison Schools
(Non-parametric Fan locally weighted regression)

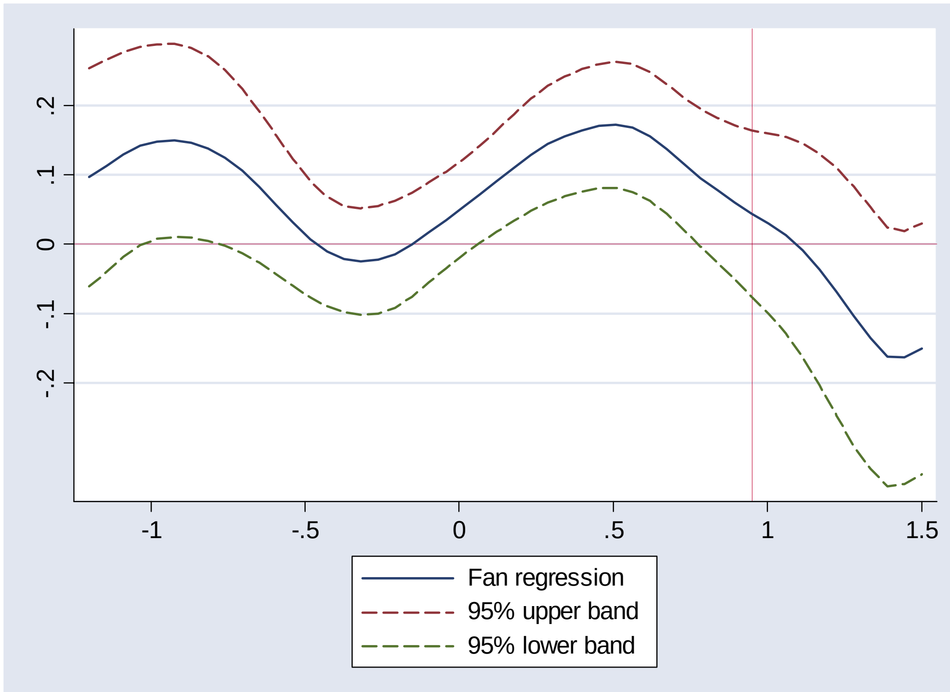

Table 1: Summary Statistics

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	2720			3254		
Number of students: Comparison	2638			3116		
Number of students: Program, Busia	1547			1843		
Number of students: Comparison, Busia	1612			1913		
Number of students: Program, Teso	1172			1411		
Number of students: Comparison, Teso	1026			1203		
Gender (1=Male)	5358	0.51	0.50	6370	0.52	0.50
Age in 2001	4937	14.3	1.6	5895	13.3	1.6
Test Score 2000	3216	0.06	0.99	--	--	--
Test Score 2001	4040	0.09	0.99	--	--	--
Test Score 2002	3404	0.05	1.01	3620	0.04	1.01
Mean School Test Score 2000	4932	0.12	0.65	5847	0.13	0.65
Attendance 2001	4798	0.79	0.41	5761	0.77	0.42
Attendance 2002	4686	0.77	0.33	5625	0.77	0.32

Panel C: **Student restricted sample**	Cohort 1, Restricted			Cohort 2, Restricted		
	Obs	Mean	Std dev	Obs	Mean	Std dev
Number of students: Program	1813			1777		
Number of students: Comparison	1908			1721		
Number of students: Busia district	2419			1869		
Number of students: Teso district	1302			1629		
Gender (1=Male)	3721	0.51	0.50	3498	0.55	0.50
Age in 2001	3721	14.2	1.5	3498	13.1	1.5
Test Score 2000	2414	0.13	0.97	--	--	--
Test Score 2001	3721	0.09	0.99	--	--	--
Test Score 2002	2803	0.11	1.01	3498	0.05	1.01
Mean School Test Score 2000	3721	0.14	0.64	3498	0.15	0.66
Attendance 2001	3582	0.86	0.35	3374	0.84	0.37
Attendance 2002	3527	0.83	0.27	3491	0.87	0.21

Notes: These statistics are for girls and boys in the sample. Dashes indicate that the data are unavailable (for instance, 2000 and 2001 exams for Cohort 2). Attendance in 2001 is from 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. The mean school test score in 2000 is for those students in the restricted sample.

The cohort 1 Baseline sample refers to all students that were registered in grade six in January 2001. The cohort 1 Restricted sample consists of 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, mean school test scores in 2000, and individual test scores in 2001.

The cohort 2 baseline sample refers to all students that were registered in grade five in January 2001. The cohort 2 restricted sample 2001 consists of 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, mean school test score in 2000, and individual test score in 2002.

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)	5.0	5.0	-0.0 (0.4)
Mother's education (years)	4.1	4.3	-0.2 (0.4)
Total children in household	6.2	5.7	0.5 (0.4)
Proportion ethnic Luhya	0.48	0.48	0.00 (0.07)
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: Standard errors in parenthesis. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Sample includes baseline students in cohort 1 and cohort 2 in 2001 in program and comparison schools in Busia district and Teso district. Data is from 2002 Student Questionnaire. The sample size is 7,401 questionnaires, 63 percent of the baseline sample.

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 students with 2001 test scores					**Panel A2: Baseline students with 2002 test scores**				
% Busia	0.82	0.77	0.05	(0.03)	% Busia	0.51	0.51	0.01	(0.04)
% Teso	0.63	0.77	-0.14***	(0.04)	% Teso	0.64	0.66	-0.01	(0.08)
Panel B1: Baseline students in schools that did not pull out					**Panel B2: Baseline students in schools that did not pull out**				
% Busia	0.96	1	-0.04	(0.04)	% Busia	0.95	1	-0.05	(0.04)
% Teso	0.88	0.87	0	(0.12)	% Teso	0.88	0.89	-0.01	(0.10)
Panel C1: Restricted Sample: Have 2000 and 2001 Test Scores and Age Data and in schools that did not pull out					**Panel C2: Restricted Sample: Have 2002 Test Scores and Age Data and in schools that did not pull out**				
% Busia	0.77	0.77	0.00	(0.05)	% Busia	0.50	0.50	0.00	(0.04)
% Teso	0.54	0.66	-0.12	(0.09)	% Teso	0.61	0.64	-0.03	(0.08)

Notes: Standard errors in parenthesis. 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 ICS exam scores (administered in early 2003), rather than district exam scores; the 2002 Teso district exams were cancelled due to the upcoming Kenyan national elections (as described in Section 2.2 of the text). Cohort 2 data for Busia district students in 2002 is based on the 2002 Busia district exams, which were administered as scheduled in late 2002, and for which students must pay a small fee (unlike the ICS exams, where were free, possibly explaining the lower attrition rate in Teso district in 2002 than in 2001).

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 Restricted sample (have 2000 and 2001 test scores, and age data, in schools that did not pull out)								
Average score in 2000, Busia	-0.08	-0.47	0.39**	0.09	0.11	-0.02	-0.17	-0.58***
			(0.19)			(0.20)	(0.11)	(0.09)
Average score in 2000, Teso	0.18	-0.35	0.53**	0.21	0.15	0.06	-0.03	-0.50***
			(0.26)			(0.17)	(0.23)	(0.14)
Panel B: Attrition from Baseline sample to those in schools that pulled out								
Average score in 2000, Busia	0.20	----	----	0.05	0.00	0.05	0.14	----
						(0.19)	(0.13)	
Average score in 2000, Teso	0.96	-0.53	1.49**	0.09	0.08	0.01	0.88**	-0.60**
			(0.45)			(0.17)	(0.37)	(0.25)

Notes: Standard errors in parenthesis. Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. Sample attrition in Panel B 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 in that category.

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 (1)	Girls and Boys Busia and Teso districts (2)
Program school	0.12**	0.17***
	(0.05)	(0.06)
Male*Program School		-0.08
		(0.05)
Male		0.34***
		(0.04)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	10470	10470
R^2	0.29	0.31
Mean of dependent variable	0.08	0.08
Panel B: Restricted Sample	Girls and Boys, Busia district, (1)	Girls and Boys, Teso district, (2)
Program school	0.20***	-0.02
	(0.07)	(0.07)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	6300	4170
R^2	0.35	0.21
Mean of dependent variable	0.09	0.06
Panel C: Restricted Sample	Girls, Busia district (1)	Boys, Busia district (2)
Program impact, Cohort 1 (in 2001)	0.29***	0.21**
	(0.10)	(0.09)
Program impact, Cohort 2 (in 2002)	0.21**	0.13
	(0.10)	(0.12)
Post-Program impact, Cohort 1 (in 2002)	0.28***	0.13
	(0.08)	(0.09)
Mean school test score, 2000	0.81***	0.78***
	(0.07)	(0.06)
Student age in program year	-0.09***	-0.14***
	(0.01)	(0.01)
% Luo ethnic group in school	0.02	-0.13
	(0.19)	(0.21)
% Teso ethnic group in school	0.65**	0.34
	(0.25)	(0.23)
% other (not Luhya, Luo, or Teso) group in school	-0.20	0.90**
	(0.49)	(0.36)
Sample Size	2990	3310
R^2	0.39	0.37
Mean of dependent variable	-0.04	0.21

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. OLS regressions, Huber robust standard errors in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized such that comparison group test scores had mean zero and standard deviation one. Indicator variables are included in all specifications 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, in schools that did not pull out of the program, and for whom we have age, and test score data.

Table 6: Within- and Between-school Inequality in Test Scores
Cohort 1 Girls, Busia District

	2000 Baseline	2001 Program year	2002 Post-program
Panel A: Program Schools			
Total, standard deviation	0.88	0.94	0.97
Within-school, standard deviation	0.59	0.66	0.67
Between-school, standard deviation	0.66	0.66	0.70
Panel B: Comparison Schools			
Total, standard deviation	0.92	0.90	0.92
Within-school, standard deviation	0.65	0.66	0.67
Between-school, standard deviation	0.65	0.61	0.63
Panel C: Difference in test score standard deviation between Program, Comparison schools			
H_0: difference=0, p-value			
Total standard deviation	0.40	0.30	0.26
Within-school standard deviation	0.02	0.72	0.98
Between-school standard deviation	0.69	0.07	0.02

Notes: Within-school standard deviation and between-school standard deviation are calculated fitting large one-way analysis-of-variance. Standard deviations are reported for girls within schools or between schools. Panel C reports p-values for the hypothesis that standard deviation in the program schools equals standard deviation in the comparison schools.

Table 7: Program Impact on School Attendance, Busia District

Panel A: Restricted Sample	Dependent variable: Student school attendance in 2001, 2002	
	Girls and boys, Busia district	Girls and boys, Teso district
	(1)	(2)
Program school	0.05**	-0.02
	(0.02)	(0.02)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	12591	8277
R^2	0.06	0.05
Mean of dependent variable	0.85	0.85

Panel B: Restricted Sample	Dependent variable: Student school attendance in 2001, 2002	
	Girls, Busia district	Boys, Busia district
	(1)	(2)
Program impact, Cohort 1 (in 2001)	0.06	0.08
	(0.04)	(0.05)
Program impact, Cohort 2 (in 2002)	0.02	-0.02
	(0.02)	(0.02)
Post-Program impact, Cohort 1 (in 2002)	0.02	0.01
	(0.03)	(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	5752	6320
R^2	0.89	0.89
Mean of dependent variable	0.826	0.86

Panel C: Teacher attendance	Dependent variable: Teacher attendance in 2002, Busia district
Program school	0.05***
	(0.02)
Mean school test score in 2000, and school ethnic composition controls	Yes
Sample Size	777
R^2	0.03
Mean of dependent variable	0.87

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. OLS regressions, Huber robust standard errors 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 in all specifications 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). Attendance data were collected during unannounced school visits.

Restricted sample includes students who were registered in sixth grade in January 2001, in schools that did not pull out of the program, and for whom we have age and test score data. 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. Attendance is coded as zero for any pupil that is absent or dropped out. Attendance is coded as missing for any pupil that died, transferred, or for whom the information was unknown. There was one 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.

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

	Cohort 1:		Cohort 2:	
Dependent Variables:	Estimated impact	Mean of dep. var.	Estimated impact	Mean of dep. var.
Panel A: Study/Work habits				
Student used a textbook at home in last week	0.06**	0.88	-0.02	0.87
	(0.03)		(0.03)	
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	0.79	0.01	0.76
	(0.04)		(0.05)	
Student went for extra coaching in last two days	0.00	0.36	-0.01	0.34
	(0.05)		(0.05)	
Teacher asked the student a question in class in last two days	-0.03	0.84	0.07	0.79
	(0.04)		(0.05)	
Amount of time did chores at home[Ψ]	0.01	2.67	0.00	2.64
	(0.05)		(0.07)	
Panel B: Educational Inputs				
Number of textbooks at home	0.02	1.9	0.23	1.8
	(0.18)		(0.16)	
Number of new textbooks this term	0.05	0.90	0.13	0.93
	(0.15)		(0.13)	
Number of new exercise books this term	0.11	2.79	0.17	3.01
	(0.16)		(0.15)	
Panel C: Attitudes towards education				
Student thinks she a "good student"	0.02	0.69	0.02	0.75
	(0.06)		(0.05)	
Student thinks that being a "good student" means "working hard"	-0.03	0.84	-0.02	0.75
	(0.03)		(0.04)	
Student thinks can be in top three in her class	0.02	0.36	-0.01	0.34
	(0.05)		(0.05)	
Student prefers school to other activities (index)[ΨΨ]	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 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 a program school indicator, 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 (not Luhya, Luo, or Teso) ethnic group in school (analogous to Table 5, Panel C). Restricted sample includes students who were registered in grade 6 in January 2001, in schools that did not pull out of the program, and for whom we have age, attendance, and test score data. The sample size varies from 700-850 observations, depending on the extent of missing data in the dependent variable.

[Ψ] Household chores include fishing, washing clothes, working on the farm and shopping at the market. Time doing chores included "never", "half an hour", "one hour", "two hours", "three hours", and "more than three hours" (coded 0-5 with 5 as most time).

[ΨΨ] 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. These capture aspects of student "intrinsic motivation". 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: Impact of Winning the Scholarship in 2002, Regression Discontinuity Estimates, Busia District Girls (Cohort 1 Restricted Sample)

	Dependent variable:			
	2002 School participation	Thinks she is a "good student"	Prefers school to other activities	2002 test score
	OLS	Probit	OLS	OLS
	(1)	(2)	(3)	(4)
Program school * 2001 test score above winning threshold	0.20***	0.05	-0.03	-0.20
	(0.06)	(0.15)	(0.07)	(0.19)
Program school	-0.02	-0.03	0.02	0.06
	(0.03)	(0.07)	(0.02)	(0.11)
2001 test score above winning threshold	-0.10*	-0.07	0.04	-0.14
	(0.05)	(0.13)	(0.05)	(0.14)
2001 test score polynomial controls	Yes	Yes	Yes	Yes
Age, school ethnic controls	Yes	Yes	Yes	Yes
Sample Size	1122	723	723	723
R^2	0.09	0.02	0.04	0.58
Mean of dependent variable	0.83	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 in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized such that comparison group test scores had mean zero and standard deviation one. The 2001 test score winning threshold in Busia district was 0.99 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, and test scores in 2001 in Busia district. In Table 9, columns (2) – (4), 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.

Table 10: Cost-effectiveness of Various Kenya Primary School Interventions

Project (authors)	Average test score gain, Years 1-2 (1)	Cost / pupil (2)	Cost / pupil per 0.1 s.d. gain (3)	Cost / pupil per 0.1 s.d. gain, adjustment for deadweight loss (4)	Cost / pupil per 0.1 s.d. gain, adjustment for deadweight loss and transfers (5)
Girls scholarship program					
Busia and Teso Districts	0.12 s.d.	\$4.24	\$3.53	\$4.94	\$1.41
Busia District	0.20 s.d.	\$3.55	\$1.77	\$2.48	\$0.71
Teacher incentives (Glewwe et al. 2003)	0.07 s.d.	\$2.39	\$3.41	\$4.77	\$1.36
Textbook provision (Glewwe et al. 1997)	0.04 s.d.	\$1.50	\$4.01	\$5.61	\$5.61
Deworming project (Miguel and Kremer 2004)	≈ 0	\$1.46	+∞	+∞	+∞
Flip chart provision (Glewwe et al. 2004)	≈ 0	\$1.25	+∞	+∞	+∞

Notes: All costs are in nominal US\$ at the time the particular program was carried out (all programs were conducted between 1996 and 2002). The deadweight loss and transfers adjustments are described in section 6 of the text. Column 4 is referred to as “education budget cost effectiveness” in the text and column 5 is referred to as “social cost effectiveness”.

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

Time	**Activity**
2000	
November	Fifth grade students in cohort 1 take district exams (these are baseline scores in the econometric analysis).
2001	
March	Announced Girls' Scholarship Program to Head Teachers in all treatment schools. Head Teachers disseminate information to parents and students.
June	Lightning strikes school in Teso (Korisai P.S.)
September – October	NGO holds Parent-Teacher meetings in all schools to remind parents and students of the program and upcoming tests.
September – October	Field officers perform unannounced school visits to collect attendance data.
November	Sixth grade students in cohort 1 take district exams (these measure the impact of the program).
2002	
January	NGO holds school assemblies to announce the first round of winners and give scholarships.
January – March	Field officers perform unannounced visits to schools to collect attendance data.
February – June	Field officers administer the student survey to all fifth, sixth and seventh grader students.
November	District exams are administered in Busia district. For sixth grade Busia students in cohort 2, these exams are used to determine second cohort of scholarship winners. Teso district exams are canceled due to Kenyan national elections.
2003	
February	NGO administers standardized exams in both Busia and Teso districts. These exams are used to determine the second round of scholarship winners, among Teso students in cohort 2 who were in sixth grade in 2002.

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

	English	Geography / history	Math	Science	Swahili
	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)
Program impact, Cohort 1 (in 2001)	0.27***	0.26***	0.25	0.26**	0.12
	(0.08)	(0.10)	(0.15)	(0.11)	(0.08)
Program impact, Cohort 2 (in 2002)	0.07	0.19**	0.28**	0.30***	0.04
	(0.10)	(0.10)	(0.13)	(0.10)	(0.11)
Post-Program impact, Cohort 1 (in 2002)	0.17***	0.12	0.28**	0.25**	0.24**
	(0.07)	(0.08)	(0.11)	(0.10)	(0.10)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes	Yes	Yes	Yes
Sample Size	2981	2981	2981	2981	2981
R^2	0.37	0.35	0.18	0.25	0.22
Mean of dependent variable	0.02	-0.13	-0.01	-0.11	0.11
F-test, impacts equal across cohorts (p-value)	0.14	0.31	0.98	0.90	0.30

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. OLS regressions, Huber robust standard errors in parenthesis. Disturbance terms are allowed to be correlated across observations in the same school, but not across schools. Test scores were normalized such that comparison test scores had mean zero and standard deviation 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 in all specifications for Cohort 1 in 2001, Cohort 1 in 2002, and Cohort 2 in 2002 (coefficient estimates not shown). Restricted sample includes cohort 1 and cohort 2 students who were registered in grades six and five, respectively, in January 2001, in schools that did not pull out of the program, and for whom we have age, and test score data .