

Incentives to Learn

Michael Kremer^{*}

Edward Miguel^{**}

Rebecca Thornton^{***}

May 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 received a cash grant and had school fees paid. Girls eligible for the scholarship showed significant gains in academic exam scores (average gain 0.15 standard deviations). There was considerable sample attrition and no significant program impacts in the smaller of the two program districts, but in the other district girls showed large gains (average gain 0.22-0.27 s.d.), and these gains persisted one full year 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. Both student and teacher school attendance increased in the program schools.

* Department of Economics, Harvard University, The Brookings Institution, and NBER. Department of Economics, Littauer 207, Harvard University, Cambridge, MA 02138, USA; 1 (617) 495-9145, mkremer@fas.harvard.edu.

** Department of Economics, University of California, Berkeley and NBER. Department of Economics, 549 Evans Hall #3880, University of California, Berkeley, CA 94720-3880, USA; 1 (510) 642-7162, emiguel@econ.berkeley.edu.

*** Department of Economics, Harvard University, Littauer 207, Cambridge, MA 02138, USA; rlthornt@fas.harvard.edu.

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, Maureen Wechuli, and the PSDP field staff and data group, without whom the project would not have been possible. Gratitude is also extended to the teachers and school children of Busia for participating in the study. George Akerlof, David Card, Brian Jacob, Victor Lavy, Antonio Rangel, Joel Sobel, and Doug Staiger have provided valuable comments. We are grateful for financial support from the World Bank, the Young Green Foundation, and the MacArthur Foundation. All errors are our own.

1. Introduction

While most education research focuses on the effects of material inputs, class size, or school organization, the most important input in the education production function may be children's study effort. Yet the impact of performance incentives to boost student effort has been relatively unexplored in both economics and education. There are a number of reasons to think that children's study effort may often be sub-optimally low in the absence of performance incentives. For one, even in the presence of large future returns to education in the labor market, individual study effort may be too low in a lifetime sense since adolescents typically have far higher time discount rates than adults (Greene 1986, Nurmi 1991). In a second, related point, agents with time inconsistent preference, such as hyperbolic discounting, may under-invest in activities like studying which have immediate costs and mainly long-run payoffs.

This paper examines the impact of the Girls Scholarship Program (GSP), introduced in rural Kenyan primary schools in 2001. The project 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). Scholarship program schools were randomly selected, allowing us to attribute differences in educational outcomes between the program and comparison groups to the scholarship program. We employ data on student and teacher school attendance, and students' self-esteem, attitudes toward school, and study patterns to isolate the behavioral mechanisms through which program incentives impact education.

There are three main findings. First, program school girls had significantly higher test scores than comparison school girls on average, and this result is nearly identical over the two years of scholarship competition. Second, program test score impacts persisted for one full year after the competition, evidence of medium-run program benefits that are unlikely to be simply the result of cramming for a single exam, or "teaching to the test". Third, 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 initial scores, who had little chance at winning the award. The existence of externality benefits greatly increases

program cost-effectiveness, since even those not provided the incentive (i.e., boys) show large gains, and these externalities provide an additional rationale for subsidizing study effort.

Program effects are concentrated in Busia district. Problems with program implementation, local opposition to the scholarship program in some communities, and resulting higher rates of sample attrition complicate the interpretation of estimated impacts in the smaller study district (Teso district). The failure to generate academic gains in Teso district suggests that parent and community support may be a necessary condition for the success of student incentive programs, and that these programs may not be worthwhile in the absence of such support. For example, we find that teacher attendance increased in Busia district program schools but not in Teso program schools, and informal parent pressure on the school teachers to improve their performance is a plausible explanation why. Community support for the program may also increase the non-monetary utility benefits of winning the award, in terms of local social prestige, further motivating students to exert more study effort.

In terms of the underlying behavioral channels, student school attendance was significantly higher in Busia district program schools, evidence that study effort increased in response to the incentive. Girls in program schools were also somewhat more likely to use textbooks to study at home, providing 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). Social psychological factors are unlikely to be driving the test score effects: there are no statistically significant changes in students' self-expressed attitudes toward school, or toward their own academic ability.

These findings contrast with those from a related program in Kenya which provided incentives for teachers based on student test score performance (Glewwe et al., 2003). While both programs led to short-run test score increases, the teacher incentive program had no measurable effect on teacher or student attendance and no persistent effect on test scores, but it did increase the frequency of test preparation sessions – suggesting a form of “teaching to the test” rather than genuine learning.

The findings of the current paper also speak to ongoing policy debates in education circles in both industrialized and less developed countries regarding the desirability of academic merit awards. Many scholarships in the United States were merit-based historically, but over time there was a 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 still based on financial need, merit funds have grown almost 50% in the past five years (College Board 2002).

A number of studies suggest university scholarships increase enrollment (for instance, Dynarski 2003), but few examine the incentive effects of merit scholarships, and those that do find mixed impacts. Binder et al. (2002) find that while scholarship eligibility in New Mexico increased student grades, the number of credit-hours students completed decreased – suggesting that students took fewer courses to keep their grades high. Similarly, the average SAT score for Georgia’s high school seniors rose almost 40 points after the HOPE college scholarship program was introduced (Cornwell et al. 2002), but it resulted in a 12% decrease in full course-load completion, 2% average reduction in completed college credits, 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.

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

The program evaluated in Israel differs from the one evaluated in this paper in several important ways. First, due to political and logistical concerns, 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 is much larger, containing more than three times as many schools as Angrist and Lavy – 127 versus

40. 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 schools were similar. Third, in addition to test score outcomes, we collected data on student school attendance, teacher attendance, and a range of student attitudes and behaviors 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 impact of increased student effort.¹

The current study also contributes to the diverse literature on the impacts of extrinsic versus intrinsic motivation. In accordance with standard economic models, incentives should increase individual study effort. An alternative theory, from psychology, asserts that extrinsic rewards may interfere with intrinsic motivation and could thus actually reduce effort. A weaker version of this second view is that incentives may lead to greater performance over the short-run, but could lead to negative effects over the longer-term, and especially when the incentive is removed, by diminishing intrinsic motivation.

Early experimental psychology research in education supported the idea that reward-based incentives lead to increased effort in students, and this became a foundation for “positive reinforcement” teaching methodologies (Skinner 1961). Laboratory research conducted in the 1970’s studied behavior before and after individuals received extrinsic motivational rewards and found evidence that 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 and Pierce (1994) and 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 Deci et al.

¹ 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, Bettinger et al (2002) show that a Colombian program that made a certain minimum level of academic performance a requirement for receiving a high school voucher led to academic gains there.

(1999) conclude that there are usually negative effects of rewards on task interest and satisfaction in a similar meta-study. The current study differs from much of the work in psychology by estimating the impacts of external rewards in a real-world context rather than in laboratory experiments. We ultimately find evidence consistent with the first view: the introduction of the scholarship did not reduce the performance of either eligible (girl) or ineligible (boys) students in the program schools, up to a full year after the scholarship competition, nor did it reduce their stated enjoyment of school activities.

The remainder of the paper proceeds as follows. Section 2 provides background 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 set and sample attrition, and Section 5 presents the estimation strategy and the empirical results. 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.² Admission to secondary school depends on performance on the official Kenya Certificate of Primary Education (KCPE) exam in Grade 8, and students take the 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 violence during national election years. End-of-year exams are standardized in each district and test students in five subjects, English, Math, Swahili, Science, and Geography/History. Students must pay a fee to take the exam – we discuss implications of this fee below. Kenyan district education offices have a well-established system of exam supervision, with exam proctors

² For instance, girls in the baseline sample (in the comparison schools) had a dropout rate of 9% from January 2001 through early 2002 versus 7% for boys. Drop-out rates were slightly lower in program schools (not shown).

(“invigilators”) from outside the school monitoring all exams. 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), ICS Africa, in Busia district and Teso district, two rural districts in western Kenya. 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 observable socio-economic characteristics, such as 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 GSP) 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. 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. Historically, Tesos in this area were educationally disadvantaged relative to Luhyas, with far fewer Teso than Luhya secondary school graduates, for example, and it has been claimed that indigenous religious beliefs related to traditional taboos and witchcraft remain stronger in Teso district than in Busia (Government of Kenya 1986).

During both years of the scholarship program (2001 and 2002), students had to pay school fees set by their local school committee. These school fees are used for non-teacher costs including chalk, textbooks, and classroom repair. 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 the policy was unclear on whether parents could impose “voluntary fees” to cover school inputs. As a result, school committees in this area continued collecting fees, which amounted to approximately US\$6.40 (500 KSh)³ per family each year on average. In practice, while these fees set a benchmark for bargaining between headmasters and parents, most parents did not pay the full fee. In

³ One US dollar was worth 78.5 Kenyan shillings in January 2002 (Central Bank of Kenya 2002).

addition to the per family school fee, there are also fees for particular activities, such as taking standardized exams (as noted above), and families must pay the full cost of books, supplies, and school uniforms (the average uniform costs US\$6.40 dollars) for their children.

Mwai Kibaki became president of Kenya following December 2002 elections, and put a policy of free primary education into place in early 2003. Unlike previous announcements of free education during the Moi period, this policy was 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. Note that 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.

2.2 Project Description and Timeline

The Girls Scholarship Program was announced to the sample schools in March 2001. Out of 127 sample primary schools, half were randomly chosen to be invited to participate in the program. The randomization was stratified by administrative divisions (there are eight such divisions in Busia and Teso districts), and along participation in a past NGO assistance program, which had provided classroom flip charts.⁴ 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 adds a component of immediate gratification to the usual benefits of 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). In each year, the award consisted of: (1) a grant of US\$12.8 (1000 Kenyan shillings, KSh) paid to the girl's family; (2) a grant of US\$6.4 (500 KSh) intended for the winner's school fees and paid directly to her school; and (3) public recognition at a school awards assembly organized by the NGO.

⁴ All GSP schools had previously participated in the flip chart program, and are a subset of that sample. The flip chart program schools were chosen since they had not been recipients of several previous NGO school assistance programs. There is no evidence that the flip chart program affected test scores. These schools are representative of local schools along most dimensions – see Glewwe et al. (2004) for details on sample characteristics.

Note that there may be benefits for winners' siblings as well, due to the income transfer, because primary school fees were levied per household, rather than per student, so the cost of schooling declined for them as well, and any within household learning spillovers (we plan to estimate sibling impacts in future research). 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, and this may have led to increased cooperation with school headmasters.⁵ Some of these funds may also have benefited teachers, if they were used to improve the staff room, for example, although these transfers were generally small.

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 and those registered for grade 5 in January 2001 made up the second cohort (and they competed for the award in 2002). The NGO restricted eligibility to those 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 in fact in all only 5% of sample students transferred schools between January 2001 and January 2002, and incoming transfer rates were 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.⁶ In March 2001, the NGO held meetings with the headmasters of schools chosen for the program to inform them of program rules and to give each school community the choice to participate. Headmasters were asked to relay information about the program via a school assembly. Because of variation in the extent to which headmasters effectively disseminated information about the scholarship, 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

⁵ Although mandatory school fees were abolished in 2003, as described above, the NGO continued to pay grant money directly to schools with scholarship winners in 2003.

⁶ A detailed project timeline is presented in Appendix Table A. Unfortunately, there is incomplete 2000 baseline exam data for cohort 2 (when they were in grade 4), and thus baseline comparisons focus on cohort 1.

program rules in advance of November 2001 district exams. After these meetings, enumerators began collecting school attendance data during unannounced visits.

The 2002 student survey indicates that girls had excellent knowledge of the program: 88% of cohort 1 and 2 girls in Busia district claimed to have heard of the program, and knowledge levels were only slightly worse in Teso district (85%). Girls had somewhat better knowledge about program rules governing eligibility and winning than boys: Busia girls were 8 percentage points more likely than boys to know that “only girls were eligible for the scholarship” (86% to 78%) and 7 percentage points more likely to know exactly how many scholarships would be awarded (32% to 25%), and patterns are again similar in Teso district (not shown). Girls in Busia were only slightly more likely than boys to report that their parents had mentioned the program to them at home (72% to 70%).

In June 2001, a tragic incident occurred in which lightning struck a primary school in Teso district (Korisai Primary School), severely damaging the school, killing seven students and injuring 27 others. Because ICS had been involved with an assistance program in that school, and due to certain strange coincidences (the names of certain lightning victims were the same as the names of ICS staff members who had recently visited the 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 district school located near the Teso 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. We discuss implications of this tragedy for econometric inference in Section 4.2 below.

Students took Grade 6 district exams in November 2001, and each district gave a separate exam. Scholarship winners were chosen based on their total score across all five subject tests. The NGO then awarded scholarships to the highest scoring 15 percent of Grade 6 girls in the program schools in each district (this amounted to 110 girls in Busia district and 90 in Teso district). Schools varied considerably

in the number of winners, but nearly 60% of program schools (36 of 63 schools) had at least one winner in 2001. Among schools with at least one 2001 winner, there was an average of five winners per school.

The NGO held school assemblies – for students, teachers, parents and local officials – in January 2002 to announce and publicly recognize the 2001 winners. Each winner was awarded a certificate, parents received the US\$12.8 (1000 Ksh) cash grant, and the school received US\$6.4 (500 Ksh) to cover school fees. The community was reminded that the program would continue for another year. Parents of winning girls were instructed that the grant should be used to purchase school materials for the winning girl, such as school uniforms, textbooks, exercise books, and pens.⁷

During the 2002 academic year, the NGO returned regularly to both program and comparison schools to conduct unannounced attendance checks and administer questionnaires to students in Grades 5-7, to collect information on their schooling effort, habits, and attitudes (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 the possibility of disruptions due to the upcoming 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, over 70% of the program schools (44 of 63 schools) had at least one winner, an increase over 2001.

3. A Simple Model of Incentives and Study Effort

A stylized economic framework helps to illustrate potential impacts of merit awards on the academic test score, $TEST_{ist}$, of student i in school s in period t . Study effort, E_{ist} , may take various forms, some of which are relatively easy to observe and measure, including improved school attendance, and others that are more problematic to measure in practice, such as paying better attention in class. In addition to individual study effort, test scores may also be a function of: the average study effort of other children in

⁷ Informal 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.

the class, \bar{E}_{st} , since it may be easier to learn when classmates are also serious about their studies, a point developed in Lazear (2001); teacher effort, E_{st}^T , which can take the form of improved teacher attendance or greater effort preparing lessons; as well as a function of the child's current academic ability level (or "human capital"), $H_{is,t-1}$, which is a function of past effort exerted by the child herself, by her classmates, and her teacher, as well as a function of the child's innate ability (H_{iso}).⁸ We ignore other important inputs into educational production (e.g., textbooks, blackboards, and chalk) for simplicity. Thus, current child ability can be expressed as:

$$(1) \quad H_{ist} = F(E_{ist}, \bar{E}_{st}, E_{st}^T, H_{is,t-1}).$$

In practice, test scores proxy for individual ability:

$$(2) \quad TEST_{ist} = TEST(H_{ist}) + \varepsilon_{ist},$$

where $TEST$ is monotonically increasing in ability, and ε_{ist} is a white noise error term reflecting random variation in measured test score outcomes. Individual learning is increasing in each argument of the function F . The effort of children and their teachers, and of children and their classmates may be either complements (formally, $F_{12} > 0$ and $F_{13} > 0$, respectively) or substitutes, and we do not impose either in the model. Similarly, own effort and ability, and the effort and ability of others in the classroom, may be either complements or substitutes contemporaneously, and own effort at one point in time may complement or substitute effort at other points in time (working through the ability term). If effort and ability are complements, then a one-time increase in study effort that boosts student ability – through the scholarship incentive, for instance – could lead to persistent increases in both school performance and study effort.⁹ Complementarities also open up the possibility of multiple classroom equilibria, some with high levels of effort by all and others with a poor learning environment.

⁸ We need not impose the restriction that H_{ist} is a sufficient statistic for previous educational inputs, although this simplifies the expressions.

⁹ 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.

The Girls Scholarship Project that we study directly affected incentives to exert effort. Effort increases the perceived probability, P , that an individual will win the scholarship. Winning the scholarship has total monetary and non-monetary value $V_s \geq 0$, where 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 the individual's test score and of assignment to a program school, T_s , which takes on a value of one for program ("treatment") schools. The probability of winning the scholarship is zero for all students in comparison schools, as well as for all boys (and for girls in grades other than Grade 6) in the scholarship schools. Independent of the program, ability leads to perceived time discounted future wage and non-wage benefits B , where these benefits are concave increasing in ability. The cost of exerting study effort is the convex increasing function $C(E_{ist})$. Thus the student's maximization problem over effort takes the form:

$$(3) \quad \underset{E_{ist}}{Max} \{ P(T_s, TEST(H_{ist})) \cdot V_s + B(H_{ist}) - C(E_{ist}) \}$$

Under fairly general conditions, the introduction of the award leads to greater school effort among those eligible for the award, and among those who will be eligible in future years (since they seek to increase their academic ability to boost future chances of winning). The award can also lead to persistent test gains, since a one-time increase in effort raises future ability.

A simple extension implies that teachers in program schools could also exert more effort than teachers in comparison schools. If teachers face a maximization problem similar to equation 3, in which they experience benefits 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, too, and this complementarity could generate greater teacher effort. Greater 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. It is possible that these social sanctions could differ across communities as a function of

local parent interest in, or support for, the program (for instance, across Busia and Teso districts), and in this case the merit award would generate larger gains where parents are more supportive.

This framework illuminates 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 baseline academic ability) might benefit from the program, through several possible channels. First, greater effort by classmates (\bar{E}_{st}) could improve the classroom learning environment and boost scores through a peer effect. Second, they could directly benefit from increased teacher effort (E_{st}^T), 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 teachers' and classmates' effort in education production, even children without any direct program incentives could find it optimal to exert additional effort themselves, again boosting their scores. For example, studying becomes more attractive relative to daydreaming in class if the teacher is present in the classroom and one's classmates are also studying and learning. Fourth, if individuals experience utility benefits from their relative ranking in the entire class (not modeled above), 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. Finally, if the merit award boosts school attendance or enrollment among girls striving for the award, and if adolescent boys prefer to attend school when more adolescent girls are at school, then the program would increase their school participation as well.

In the empirical work that follows we focus on reduced form estimation of this model, in other words, the impact of the incentive program on test scores. We also estimate program impacts on multiple possible channels linking behavior to test scores – in particular, measures of student and teacher effort, as well as other factors (e.g., student attitudes toward school) that are not explicitly modeled above – to better understand the mechanisms underlying the reduced form estimates. We do not attempt to definitively disentangle the relative contributions of each channel in the overall program impact, as a

number of possibly important channels (i.e., paying more attention in class) are difficult to measure in practice, and since this would likely require strong, untestable structural econometric assumptions.

4. Data and Estimation

4.1 The Dataset

The test score data were collected 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 cohort 1 and cohort 2 students is called the *baseline sample* (Table 1, Panel B). In the main analysis, we focus mainly on students with complete age, school attendance, and gender information, in schools that did not pull out of the program and for which we have 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 are typically somewhat below average in terms of academic achievement, as discussed below.

Attendance data are based on four unannounced checks, one conducted during September to October 2001, and one in each of the three terms of the 2002 academic year. These were collected by NGO enumerators and they recorded baseline students actually in school on the day of the unannounced check as “present”. Attendance rates are below 80% for the baseline sample and a bit over 80% for the restricted sample (Table 1, Panels B and C). These data from unannounced checks are a substantial improvement over data from official school attendance registers, which are often thought to be unreliable in less developed countries.

Household characteristics are similar across program and comparison schools (Table 2): there are no significant differences in terms of parent education, number of siblings, 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.¹⁰ Further evidence is provided by comparing the 2000 (baseline) test score distributions, which are nearly identical graphically for cohort 1 girls in Busia (Figure 2). More formally, we cannot reject the hypothesis that average baseline test scores are the same across program and comparison schools for students in the restricted sample, as discussed below.

Another estimation concern is the possibility of cheating 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 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. There are also program impacts on several objective measures of student and teacher effort – most importantly, school attendance measured during unannounced enumerator school visits.¹¹

4.2 Sample Attrition

Teso district primary schools had higher rates of sample attrition than Busia schools in 2001. The gap in attrition across program versus comparison schools was also greater in Teso district, and students who attrited from the sample in Teso appear to be systematically better students than those who participated in the program. These patterns all complicate causal inference in Teso district.

It seems likely that the high attrition in Teso district, particularly in schools near the lightning strike, was due in part to the history of misunderstanding with NGOs in that area, which has persisted despite ICS attempts to address the problem through community meetings. The tragic lightning incident

¹⁰ This comparison in Table 2 relies on the assumption that the household characteristics (i.e., parent education, fertility, and asset ownership) were not directly affected by the scholarship program by mid-2002, which seems reasonable. There is no analogous baseline household survey data.

¹¹ 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 for 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 results.

was literally the spark that set off additional hostility, leading several schools to immediately pull out of the program. As further evidence of mistrust toward NGOs, one girl in Teso district who won the ICS scholarship in 2001 refused the scholarship award (see Figure 1).

As discussed above, six of the 127 schools invited to participate refused, leaving 121 schools. Six additional schools (four in Teso district and two in Busia) with incomplete 2000, 2001, or 2002 exam scores, or missing demographic and ethnicity data were also dropped, leaving 115 schools and 6983 students in the restricted sample (students in the program schools account for 50 percent of this sample).

There is a large and statistically significant difference in attrition across program and comparison schools in Teso district schools, but much less so in Busia. Among baseline sample cohort 1 students in Teso, 63 percent of scholarship school students took the 2001 exam, while the rate for comparison school students is 77 percent (Table 3, Panel B1). Sample attrition was also more widespread in Teso district schools than in Busia district. Thus while only four percent of students in Busia district were in schools that pulled out of the program, fully 12 percent of students in Teso were in schools that left the program (Table 3, Panels C1 and C2).

There is much less evidence for differential attrition across program and comparison schools in Busia district. Among cohort 1 students, 82 percent of baseline students in Busia scholarship schools and 77 percent in comparison schools took the 2001 exam. Thus there is a small, positive but insignificant point estimate of 0.04 on the difference between the proportion taking the 2001 exam between scholarship and comparison schools (Table 3, Panel B1). Among cohort 2 students the difference is even smaller: 52 percent of scholarship school students and 51 percent of comparison students in Busia district took the 2002 exam (Panel B2). There is more overall attrition by 2002 as some students had dropped out of school, transferred to other schools, or decided not to take the district exam.

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

Differential attrition between program and comparison schools in Teso district is smaller among cohort 2 students than cohort 1. To 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, although it is difficult to test this given unobservable dimensions of student academic quality.¹² Students who did not take the 2001 exam (“attritors”) tended to be somewhat lower achieving students at baseline in both Busia and Teso districts (Table 4, Panel B). Examining the differences in 2000 baseline test scores between attritors and non-attritors shows that Busia program school students who did not take the 2001 exams scored 0.07 standard deviations lower at baseline than those who did take the 2001 exams (column (i)-(iv)). The difference is 0.56 standard deviations in Busia comparison schools (column (ii)-(v)), providing evidence that a greater proportion of low performing students attrited from the comparison school sample than from program school sample. This suggests that program impact estimates are likely to be lower bounds on true effects. Below we discuss the construction of bounded estimates that take into account potential attrition bias.

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 1.48 standard deviations higher in 2000 on average (Table 4, Panel A). 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. In what

¹² 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 for this.

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, and where program effects were larger, though we also present the main results for the full sample.

4.3 Estimation Strategy

The main estimation equation is:

$$(4) \quad TEST_{ist} = Z_{ist}' \beta + (Z_{ist} * T_s)' \gamma + X_{ist}' \delta + \mu_s + \varepsilon_{ist}$$

$TEST_{ist}$ is the test score for student i in school s and 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. X_{ij} is a vector that includes other explanatory variables, including student age, the average baseline school 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 in the same school. The disturbance terms consist of μ_s , a school effect perhaps capturing common neighborhood or headmaster characteristics, and an idiosyncratic term, ε_{ist} , which may capture unobserved student ability or shocks. A non-parametric locally weighted regression technique following Fan (1992) allows us to estimate average program impacts across individuals with different baseline scores. We use a similar empirical approaches to estimate the impact of the program on the various behavioral channels (e.g., school attendance) potentially linking the program to test scores.

5. Main Empirical Results

5.1 Academic Test Score Impacts

The scholarship program raised test scores by 0.11 standard deviations (standard error 0.05) overall among boys and girls in 2001 and 2002, pooling Busia and Teso districts and students of both cohorts (Table 5, Panel A), and this effect is statistically significant at 95 percent confidence. The estimated impact of the program is larger for girls, as expected, with a sizeable average gain of 0.15 standard deviations (standard error 0.06, statistically significant at 99 percent confidence) overall in both Busia and

Teso, while the average effect for all boys is 0.08 (not statistically significant). The overall effect for girls and boys together is considerably larger for Busia district (0.19, standard error 0.07, Panel B) than for Teso (-0.04, standard error 0.07).

We next separately estimate effects for girls and boys across cohorts and years. The program effect for girls competing for the scholarship is 0.27 standard deviations in the restricted sample of girls in cohort 1 in Busia (competing in 2001), and 0.22 for cohort 2 in 2001 (Table 5, Panel C, regression 1), and in both cases the effects are significantly different than zero at 95 percent confidence. These are large impacts: to illustrate from previous research in 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 estimated program gains roughly correspond to an additional 0.22-0.27 grades of primary school learning.

To address non-random sample attrition, we also construct non-parametric bounds on the main program treatment effects following Lee (2002), and these bounds are reasonably tight, 0.23-0.31 standard deviations in the restricted sample of cohort 1 Busia girls in 2001 – perhaps not surprising given that there is minimal differential attrition across program and comparison schools (Table 3). The main result for cohort 1 girls in Busia is also robust to using the change in test scores between 2000 and 2001 as the dependent variable (coefficient estimate 0.20, standard error 0.12 – not shown).

Other explanatory variables have expected effects. The baseline school average test score in 2000 is significantly positively associated with the 2001 test score (Table 5, Panel C). Being one year older decreases test scores by 0.02-0.03 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. Program impact estimates are similar if these explanatory variables are excluded, although estimates are less precise.¹³ Estimates are largely unchanged when individual

¹³ For instance, the program impact for Busia cohort 1 girls is still exactly 0.27 standard deviations in this case, but the standard error rises to 0.19, while the program impact for cohort 2 rises to 0.31 with standard error 0.18 (significantly different than zero at 90 percent confidence) – regressions not shown. 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.

demographic controls collected in the 2002 student questionnaire – including parent education and household asset ownership – are included as explanatory variables,¹⁴ and interactions of the program indicator with these household socioeconomic proxies are not statistically significant at traditional confidence levels (regressions not shown), suggesting that better-off residents do not gain disproportionately.

Test score effects in Teso district are near zero and statistically insignificant (Table 5, Panel B, regression 2). This is consistent 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 also with the hypothesized bias due to sample attrition in Teso. It remains unclear whether there might have been larger program effects had there been no lightning tragedy in 2001, or how common the problems of program implementation and sample attrition encountered in Teso would be in other settings.¹⁵

The scholarship program not only significantly raised test scores when it was first introduced in 2001, but also continued to increase scores of cohort 1 girls in program schools during 2002: the point estimate in year two is 0.24 (standard error 0.08, Table 5, Panel C regression 1) for the restricted sample, providing additional confidence that the program had lasting effects on learning, rather than simply being due to cheating or cramming for the 2001 exam.

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 needed to win the scholarship. The score distribution shifts to the right in program schools for cohort 1 in year 1 (Figure 3), cohort 1 in year 2 (Figure 4), and cohort 2 in year 2 (Figure 5).¹⁶

¹⁴ 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 would thus reduce sample size.

¹⁵ To 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 interaction term is large and negative but not statistically significant (-0.10, standard error 0.13), while the coefficient estimate on the Teso-program interaction term remains negative and marginally significant (-0.21, standard error 0.11) – suggesting that the lightning strike alone cannot account for the negative Teso effect. Thirteen schools, or approximately 10 percent of the sample, were located within 6 km of the strike. Interaction results are similar for distances between 5-7 km from the lightning strike.

¹⁶ These figures use a quartic kernel and a bandwidth of 0.7.

The largest gains appear to be near the minimum winning score threshold, suggesting that the students exerting the most additional effort are those who believed that additional effort could make the greatest difference in their chances to win the award.

Yet from these figures alone it is impossible to determine the magnitude and statistical significance of program effects at different parts of the initial test score distribution. Figure 6 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, perhaps due to extra study effort exerted to win the scholarship. 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, and also 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 (although it is difficult to reject the hypothesis that gains at the bottom of the baseline distribution are the same as gains elsewhere due to lack of statistical precision in the non-parametric analysis – not shown).¹⁷

Field reports from students and teachers in program schools describe how 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.”¹⁸ 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 monetary rewards.

¹⁷ Turning to scores by academic subject, the program had somewhat larger effects on test scores in math and science, than in English and Swahili, especially for cohort 2 (Appendix Table B). These findings are broadly similar to Case and Deaton (1999), who found that lower student-teacher ratios in South Africa affected math test scores more than language scores. Unexpectedly, impacts on Swahili scores are larger in the post-scholarship year for cohort 1; however, differences in program impacts across cohorts and years are not statistically significantly different for any of these four school subjects.

¹⁸ Source: authors’ field notes, July 15, 2002.

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 0.20 standard deviations (standard error 0.10) in the restricted sample (Table 5, Panel C, regression 2). Boys and girls in western Kenya share the same teacher and classroom, so any benefits from added classroom effort on the part of teachers is likely to spill over to the boys. In the first year, it is also possible that some boys were confused as to whether they too were eligible for the scholarship. 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 are the same in 2001 and 2002. Overall, there is no evidence the girls' scholarship program discouraged or demoralized boys.

Field reports from the program schools also suggest that a spirit of competition 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 alternative program that targeted boys rather than girls, or in which they competed against each other. Gneezy et al (2003) provide experimental laboratory data that females and males may 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 gender of the competition.

The focus so far has been on the first moment of the test score distribution, but the effect of incentives on inequality is also of theoretical interest. There is a small overall increase in test score inequality in program schools relative to comparison schools in the two years of program, but differences are minimal and not generally statistically significant (Table 6). The slight (though insignificant) increase in test score inequality in program schools is inconsistent with one particular naïve model of cheating, one in which program school teachers simply pass out test answers to children in their class.¹⁹ This would likely produce less inequality in program schools relative to comparison schools (note that the inequality

¹⁹ We thank Joel Sobel for this point.

patterns we find are not inconsistent with all models of cheating, however). The changes in inequality over time in program versus comparison schools are similarly minimal for boys (not shown).

5.2 School Participation, Attitudes and Behaviors

It is useful to explore the channels causing estimated program effects, since some mechanisms, such as cheating or increased coaching, might raise test scores without improving underlying learning. The scholarship program had a significant positive impact on 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 0.05 percentage points (0.02 standard error), which is statistically significant at 95 percent confidence, and reduces student absenteeism by approximately 30 percent. These attendance gains suggest program school students are exerting extra effort in one important dimension. There are statistically insignificant gains in Teso district (regression 2), and the estimated student school attendance gain pooling Busia district and Teso district together for girls and boys is not significant at traditional confidence levels (0.02, standard error 0.02 – 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 considerably smaller (Table 7, Panel B, regression 1). The pattern of 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 at the bottom of the test score distribution (Figure 7). The pre-program attendance gains might be due to anticipation of the future scholarship opportunity. 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. 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 why attendance gains are considerably larger in 2001 than in 2002.

Teachers in 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), and this large effect is robust to alternative specifications (not shown).²⁰

The 2002 Student Questionnaire collected information on educational inputs, study habits, and attitudes that may have affected school performance, to capture other dimensions of effort at least in part. Because the questionnaire 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.

There is a significant increase in textbook use among program girls in cohort 1 (Table 8, Panel A): girls in program schools report having used textbooks at home 6 percentage points more than in comparison schools, suggesting that the program led to more intensive studying among the girls (although there is no effect for cohort 2). However, there is no impact on the likelihood that program school students sought out extra school coaching, handed in homework, were called on by the teacher in class, or did 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 work effort 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 (Panel B), although all six point estimates in the panel are positive, providing suggestive evidence in favor of some increased parental investments. 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 0.26 additional new exercise books or textbooks at home in program schools (standard error 0.19, regression not shown).

There is no convincing evidence of 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).

²⁰ 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.

Overall, then, there is evidence of increased student and teacher effort (reflected in school attendance, and increased textbook use at home), perhaps suggestive evidence of increased parental educational investments, but no evidence of the changes in attitudes emphasized by psychologists.

5.3 Regression Discontinuity Estimates of the Impact of Winning the Scholarship

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

The coefficient estimate on the interaction term between the program school indicator variable and the indicator variable for scoring above the scholarship threshold captures the impact of winning the scholarship (coefficient f in equation 5):

$$(5) \quad Y_{is} = a + X'_{is} b + c \cdot WIN_{is}^{2001} + d \cdot T_s + f \cdot (WIN_{is}^{2001} * T_s) + \sum_{z=1}^3 \left\{ g_z (TEST_{is}^{2001})^z + h_z (TEST_{is}^{2001} * T_s)^z \right\} + u_s + e_{is}$$

There are large and marginally statistically significant school participation gains in 2002 among scholarship winners (Table 9, regression 1, point estimate 0.13, statistically significant at 90 percent confidence), suggestive evidence that paying girls' school fees has a positive impact on school participation. However, this result is not robust to changes in the set of explanatory variables, is smaller and not statistically significant when the discontinuity is estimated using only program school students (regression not shown), and the non-parametric regressions does not yield a visually sharp threshold at the

winning 2001 test score value (figure not shown), calling into question the robustness of the finding, and as a result we do not emphasize this result.

Winners are no more likely to claim that they are “good students” than non-winners (Table 9, regression 2), nor a significant impact on preferences for school relative to other activities (regression 3), or 2002 test score performance (regression 4). There is also no evidence of negative demoralization effects for non-winners in terms of either school participation or academic self-esteem among losers in program schools (captured by the coefficient estimate on the program indicator in regressions 1 and 2).

6. Conclusion

Merit-based scholarships are increasingly important in many countries, including the United States, but there is limited evidence regarding their impact on learning. 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 large positive classroom peer effects. Though there are large, significant and robust program effects in Busia district, on the order of 0.2 standard deviations on average, we do not find significant effects on test scores in Teso district. This may be due to differential sample attrition that complicates the econometric analysis in Teso district schools, or it may reflect the lower value placed on the merit award there – especially in the aftermath of the tragic lightning strike incident of early 2001. The starkly different program impact in Busia and Teso districts points to the important role that parent and community support is likely to play in the success of student incentive programs of this sort, an issue that is likely to be important for future program design.

The persistence of program effects on educational, labor market, fertility outcomes in the long-term also remains an open question. We have recently collected detailed contact information for individuals in the program and comparison schools, and in future work plan to follow up these individuals as they enter adulthood in order to estimate long-term program impacts. These findings will be important for more definitively establishing whether increased primary school learning really does have a payoff for rural Kenyan girls, and if so, what form these payoffs take in the long-run.

References

- Angrist, J. and V. Lavy (2003). "The Effect of High School Matriculation Awards: Evidence from Randomized Trials." *NBER Working Paper #9389*.
- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. (2002). "Vouchers for Private Schooling in Colombia: Evidence from Randomized Natural Experiments", *American Economic Review*, 1535-1558.
- Ashworth, K., J. Hardman, et al. (2001). "Education Maintenance Allowance: The First Year, A Qualitative Evaluation". Research Report RR257, Department for Education and Employment.
- 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.
- Cameron, J. and D. Pierce (1994). "Reinforcement, Reward and Intrinsic Motivation: A Meta-Analysis." *Review of Educational research* **64**(3): 363.
- Case, A. and A. Deaton (1999). "School Inputs and Educational Outcomes in South Africa." *Quarterly Journal of Economics*, 1047-1084.
- 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.
- 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 Ayrshire Pilot." Research Findings No. 6, Enterprise and Lifelong Learnings Report, Glasgow.
- Deci, E. L. (1971). "Effects of Externally Mediated Rewards on Intrinsic Motivation." *Journal of Personality and Social Psychology* **18**: 105-115.
- Deci, E. L., R. Koestner, et al. (1999). "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* **125**(627-668).
- Dickinson, A. M. (1989). "The Detrimental Effects of Extrinsic Reinforcement on "Intrinsic Motivation." *The Behavior Analyst* **12**: 1-15.
- Dynarski, S. (2003). "The Consequences of Merit Aid." *NBER Working Paper #9400*.

- 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." *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.
- 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.
- Nurmi, J. (1991). "How do adolescents see their future? A review of the development of future orientation and planning", *Developmental Review*, 11:1-59.
- Skinner, B. F. (1961). "Teaching Machines." *Scientific America* **November**: 91-102.
- World Bank. (2002). *World Development Indicators* (www.worldbank.org/data).

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

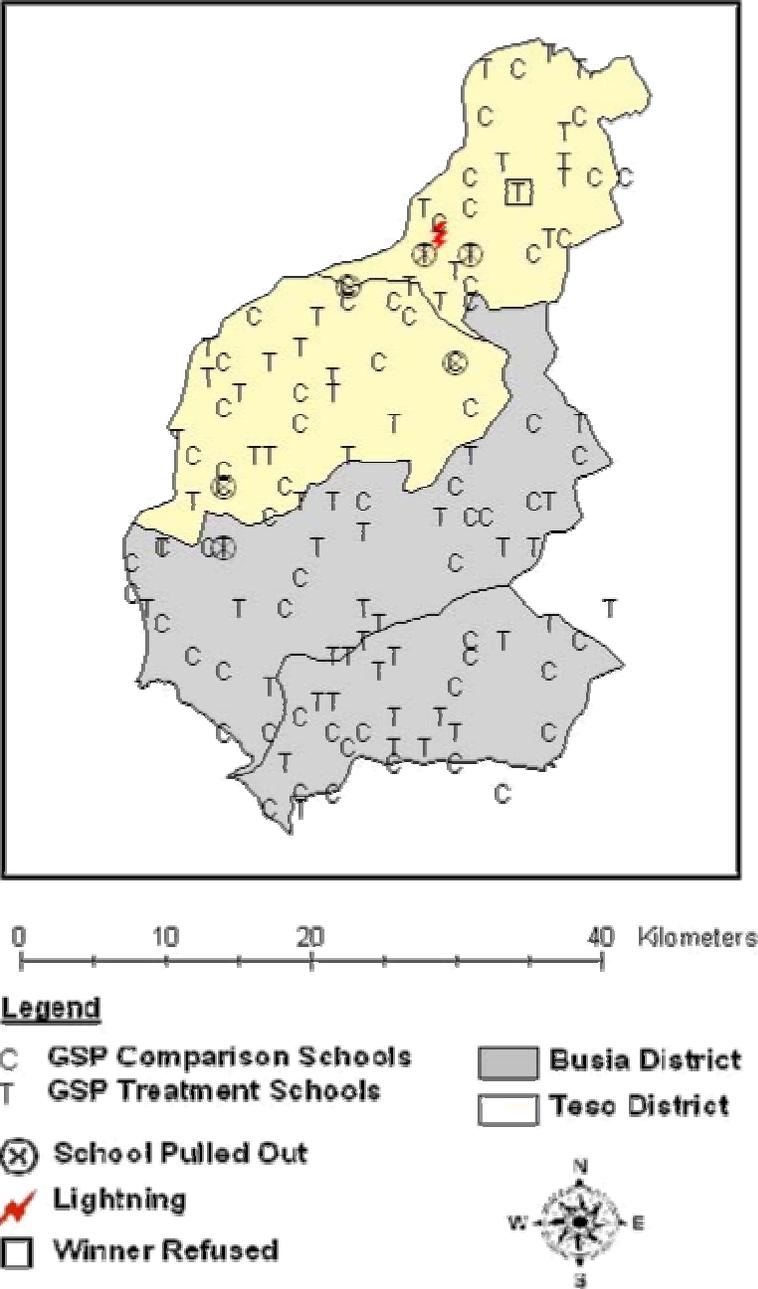
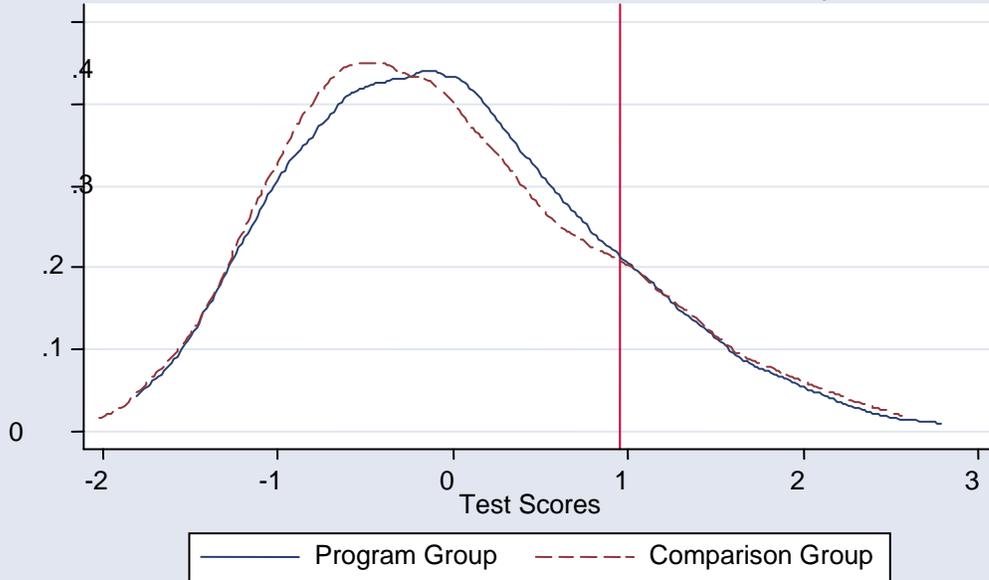


Figure 2

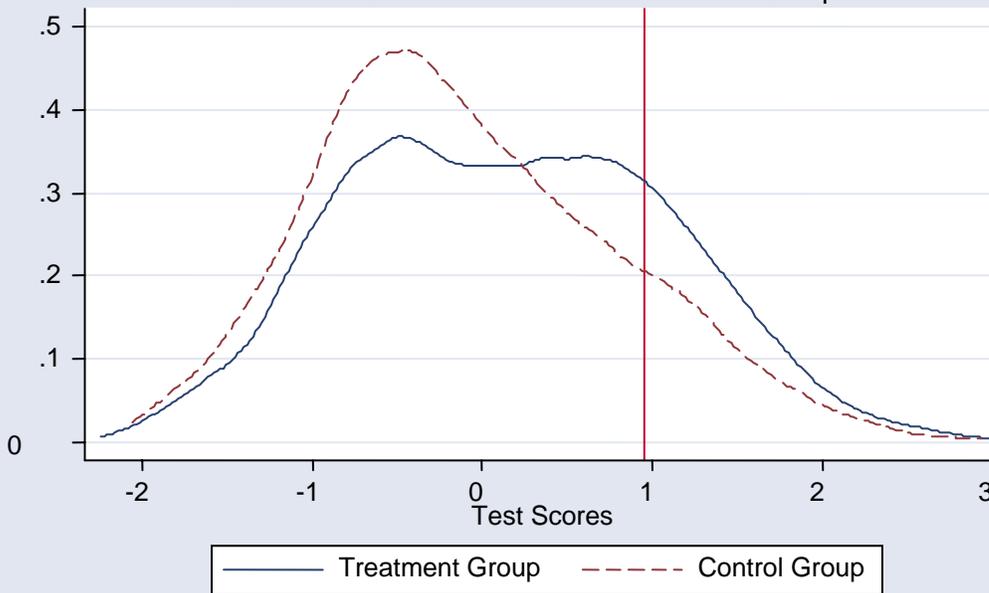
Distribution of Girl's Test Scores 2000: Busia Restricted Sample Cohort 1



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

Figure 3

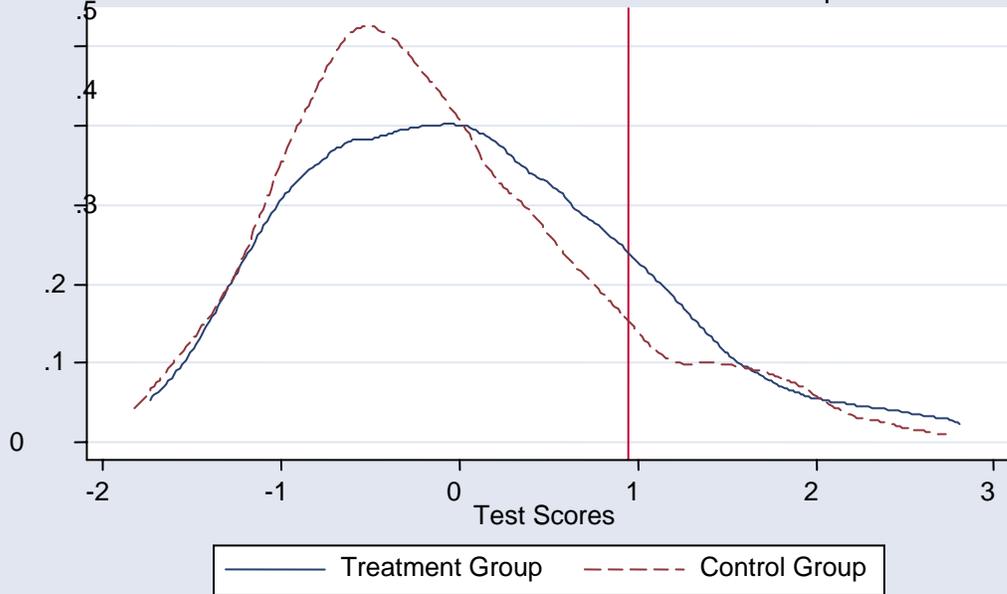
Distribution of Girl's Test Scores 2001: Busia Restricted Sample Cohort 1



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

Figure 4

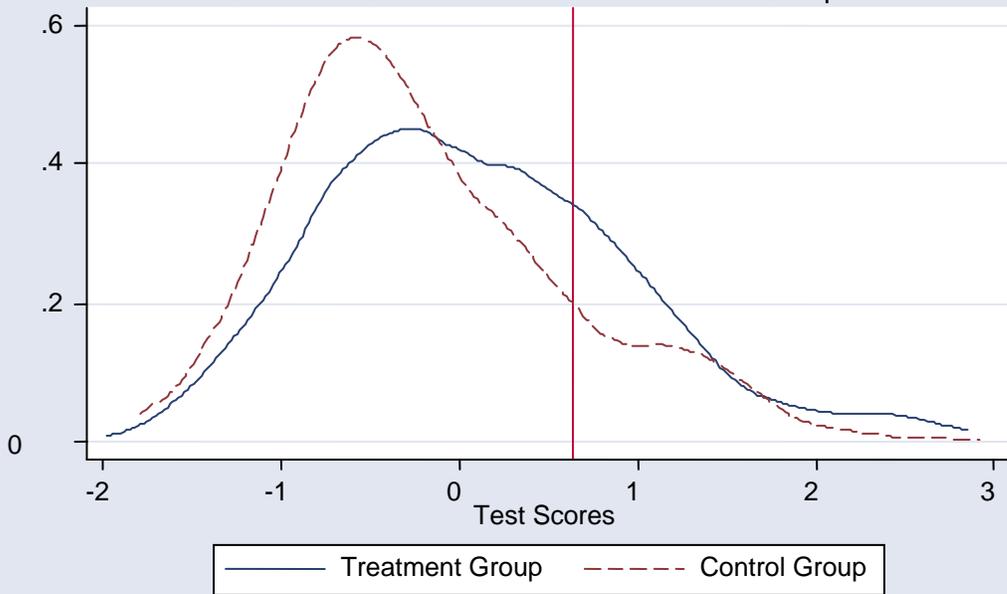
Distribution of Girl's Test Scores 2002: Busia Restricted Sample Cohort 1



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

Figure 5

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



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

Figure 6

Effect of Scholarship on 2001 Test Scores Busia Girls Cohort 1

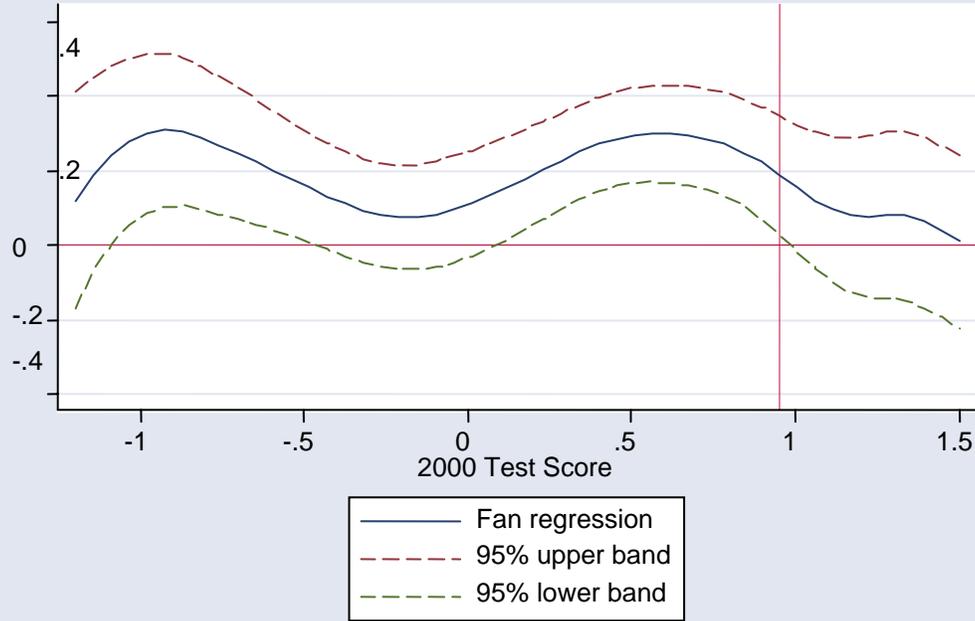


Figure 7

Effect of Scholarship on Attendance Busia Girls Cohort 1

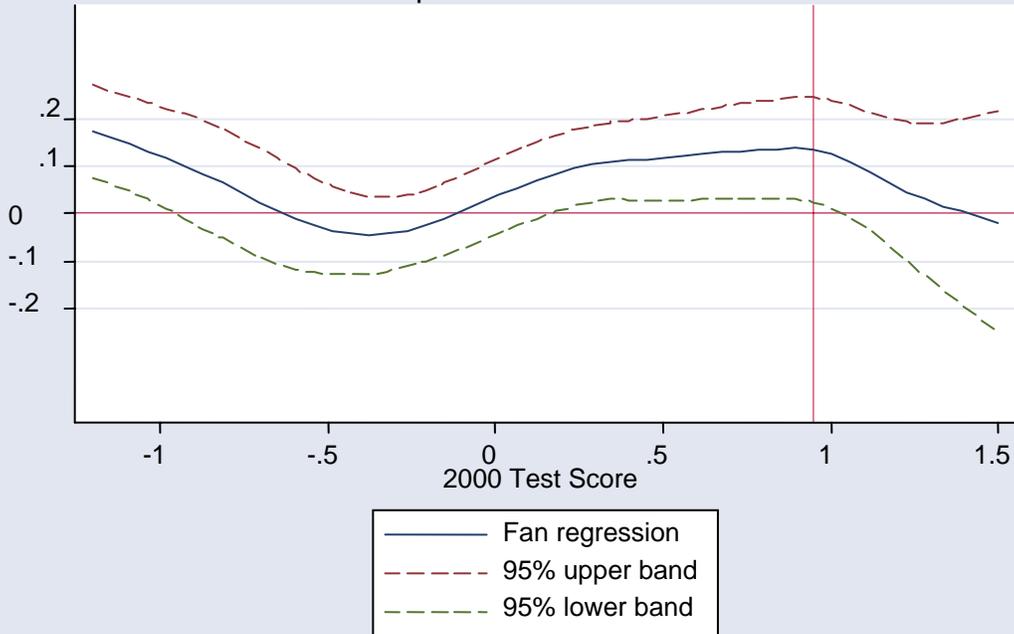


Table 1: Summary Statistics

Panel A: School characteristics						
	<u>Obs.</u>					
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		
	<u>Obs</u>	<u>Mean</u>	<u>Std dev</u>	<u>Obs</u>	<u>Mean</u>	<u>Std dev</u>
Number of students: Program	2722			3260		
Number of students: Comparison	2638			3120		
Number of students: Busia district	3162			3761		
Number of students: Teso district	2198			2619		
Gender (1=Male)	5360	0.51	0.50	6380	0.52	0.50
Age in 2001	4937	14.3	1.6	5897	13.3	1.6
Test Score 2000	3217	0.06	0.98	--	--	--
Test Score 2001	4040	0.09	0.99	--	--	--
Test Score 2002	3404	0.05	1.00	3627	0.04	1.03
Mean School Test Score 2000	5286	0.05	0.65	6265	0.07	0.66
Attendance 2001	4805	0.79	0.41	5786	0.77	0.42
Attendance 2002	4787	0.75	0.35	5742	0.75	0.34
Panel C: Student restricted sample						
	Cohort 1, Restricted			Cohort 2, Restricted		
	<u>Obs</u>	<u>Mean</u>	<u>Std dev</u>	<u>Obs</u>	<u>Mean</u>	<u>Std dev</u>
Number of students: Program	1763			1750		
Number of students: Comparison	1821			1649		
Number of students: Busia district	2366			1847		
Number of students: Teso district	1218			1552		
Gender (1=Male)	3584	0.51	0.50	3399	0.55	0.50
Age in 2001	3584	14.2	1.5	3399	13.1	1.5
Test Score 2000	2347	0.11	0.97	--	--	--
Test Score 2001	3584	0.08	0.99	--	--	--
Test Score 2002	2694	0.10	1.00	3399	0.06	1.00
Mean School Test Score 2000	3584	0.05	0.64	3399	0.06	0.65
Attendance 2001	3584	0.86	0.35	3399	0.84	0.37
Attendance 2002	3454	0.81	0.29	3387	0.87	0.21

Notes: These statistics are for girls and boys in the sample. Dashes indicate that the data are currently unavailable (for instance, 2000 and 2001 exams for Cohort 2). Attendance in 2001 is a one-time unannounced visit to schools in term 3, 2001. Attendance in 2002 is consists of three unannounced visits to schools throughout the school year.

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

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

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

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

Notes: 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 7408 questionnaires.

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

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

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).

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

	<u>Attritors</u>			<u>Non-Attritors</u>			<u>Difference</u> <u>Attritors – Non-Attritors</u>	
	Program (i)	Comparison (ii)	Difference (iii)	Program (iv)	Comparison (v)	Difference (vi)	Program (i) – (iv)	Comparison (ii) – (v)
Panel A: Attrition from Baseline sample to those in schools that pulled out								
Average score in 2000, Busia	0.19	----	----	0.05	0.00	0.05 (0.19)	0.14 (0.13)	----
Average score in 2000, Teso	0.96	-0.53	1.48** (0.45)	0.08	0.08	0.00 (0.08)	0.88** (0.37)	-0.61** (0.25)
Panel B: Attrition from Baseline sample to Restricted sample (have 2000 and 2001 test scores, age, and attendance data, in schools that did not pull out)								
Average score in 2000, Busia	0.00	-0.45	0.45** (0.19)	0.08	0.11	-0.03 (0.19)	-0.07 (0.11)	-0.56*** (0.08)
Average score in 2000, Teso	0.18	-0.20	0.37 (0.30)	0.20	0.11	0.09 (0.18)	-0.02 (0.23)	-0.30 (0.20)

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

<u>Dependent variable: Normalized test scores from 2001 and 2002</u>		
Panel A: Restricted Sample	Girls and Boys Busia and Teso districts (1)	Girls and Boys Busia and Teso districts (2)
Program school	0.11** (0.05)	0.15*** (0.06)
Male*Program School		-0.07 (0.05)
Male		0.32*** (0.04)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	10072	10072
R ²	0.30	0.33
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	0.19*** (0.07)	-0.04 (0.07)
Student age, mean school test score in 2000, and school ethnic composition controls	Yes	Yes
Sample Size	6168	3904
R ²	0.36	0.22
Mean of dependent variable	0.09	0.07
Panel C: Restricted Sample	Girls, Busia district (1)	Boys, Busia district (2)
Program impact, Cohort 1 (in 2001)	0.27*** (0.11)	0.20* (0.10)
Program impact, Cohort 2 (in 2002)	0.22** (0.10)	0.13 (0.12)
Post-Program impact, Cohort 1 (in 2002)	0.24*** (0.08)	0.13 (0.09)
Mean school test score, 2000	0.81*** (0.07)	0.78*** (0.06)
Student age in program year	-0.02*** (0.01)	-0.03*** (0.01)
% Luo ethnic group in school	0.11 (0.20)	0.11 (0.25)
% Teso ethnic group in school	0.54** (0.26)	0.29 (0.26)
% other (not Luhya, Luo, or Teso) group in school	-0.11 (0.52)	1.00** (0.43)
Sample Size	2931	3237
R ²	0.40	0.37
Mean of dependent variable	-0.04	0.19

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. 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, attendance, and test score data.

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

	<u>2000</u>	<u>2001</u>	<u>2002</u>
	<u>Baseline</u>	<u>Program year</u>	<u>Post-program</u>
<u>Panel A: Program Schools</u>			
Total (s.d.)	0.89	0.93	0.97
Within-school (s.d.)	0.59	0.67	0.67
Between-school (s.d.)	0.66	0.66	0.70
<u>Panel B: Comparison Schools</u>			
Total (s.d.)	0.91	0.90	0.92
Within-school (s.d.)	0.65	0.66	0.67
Between-school (s.d.)	0.64	0.61	0.63
<u>Panel C: Difference in test score inequality between Program, Comparison schools</u>			
H ₀ : difference=0, p-value			
Total inequality	0.39	0.30	0.26
Within-school inequality	0.03	0.72	0.98
Between-school inequality	0.69	0.07	0.02

Table 7: Program Impact on School Attendance, Busia District

		<u>Dependent variable:</u>	
		Student school attendance in 2001, 2002 Girls and boys, Busia district	Girls and boys, Teso district
		(1)	(2)
Panel A: Restricted Sample			
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		8362	5462
R ²		0.03	0.03
Mean of dependent variable		0.84	0.84
		<u>Dependent variable:</u>	
		Student school attendance in 2001, 2002 Girls, Busia district	Boys, Busia district
		(1)	(2)
Panel B: Restricted Sample			
Program impact, Cohort 1 (in 2001)		0.06 (0.04)	0.08* (0.05)
Program impact, Cohort 2 (in 2002)		0.01 (0.02)	-0.03 (0.02)
Post-Program impact, Cohort 1 (in 2002)		0.02 (0.02)	0.02 (0.03)
Pre-Program impact, Cohort 2 (in 2001)		0.10** (0.05)	0.10* (0.06)
Student age, mean school test score in 2000, and school ethnic composition controls		Yes	Yes
Sample Size		3994	4368
R ²		0.89	0.87
Mean of dependent variable		0.86	0.83
		<u>Dependent variable:</u>	
		Teacher attendance in 2002, Busia district	
Panel C: Teacher attendance			
Program school			0.05*** (0.02)
Mean school test score in 2000, and school ethnic composition controls			Yes
Sample Size			2399
R ²			0.02
Mean of dependent variable			0.86

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. 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, attendance, 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. There was one such student attendance observation in the 2001 school year, and three in 2002. The teacher attendance visits were also unannounced, and actual teacher presence at school recorded.

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

	<u>Cohort 1:</u>		<u>Cohort 2:</u>	
	Estimated impact	Mean of dep. var.	Estimated impact	Mean of dep. var.
Dependent Variables:				
Panel A: Study/Work habits				
Student used a textbook at home in last week	0.06** (0.03)	0.72	-0.03 (0.04)	0.72
Student did homework in last two days	0.01 (0.05)	0.79	0.04 (0.05)	0.79
Student handed in homework in last week	0.00 (0.04)	0.88	0.01 (0.05)	0.87
Student went for extra coaching in last two days	0.00 (0.05)	0.84	-0.01 (0.05)	0.79
Teacher asked the student a question in class in last two days	-0.03 (0.04)	0.36	0.07 (0.05)	0.35
Amount of time did chores at home	-0.01 (0.01)	1.82	0.00 (0.01)	1.80
Panel B: Educational Inputs				
Number of textbooks at home	0.01 (0.18)	1.9	0.14 (0.13)	1.8
Number of new textbooks this term	0.05 (0.15)	2.8	0.08 (0.10)	3.0
Number of new exercise books this term	0.11 (0.16)	0.9	0.18 (0.12)	0.9
Panel C: Attitudes towards education				
Student thinks she a “good student”	0.02 (0.06)	0.69	0.00 (0.05)	0.75
Student thinks that being a “good student” means “working hard”	-0.03 (0.03)	0.84	-0.02 (0.03)	0.76
Student thinks can be in top three in her class	0.03 (0.05)	0.37	-0.01 (0.05)	0.35
Student prefers school to other activities (index)	0.01 (0.01)	0.70	0.02 (0.02)	0.72

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 approximately 700-850 observations, depending on the extent of missing data in the dependent variable.

The “student prefers school to other activities” index is the average of eight binary variables indicating whether the student prefers a school activity (coded as 1) or a non-school activity (coded 0). The school activities include: doing homework, going to school early in the morning, and staying in class for extra coaching. The non-school activities include fetching water, playing games or sports, looking after livestock, cooking meals, cleaning the house, or doing work on the farm. Household chores include fishing, washing clothes, working on the farm and shopping at the market. Time doing chores included “half an hour”, “one hour”, “two hours”, “three hours”, “more than three hours”, or “never” (coded 1-4 with 4 as most time).

Table 9: Impact of Winning the Scholarship in 2002, Regression Discontinuity Estimates, Busia District Girls (Cohort 1 Restricted Sample)

	Dependent variable:			
	2002 School participation OLS	Thinks she is a "good student" Probit (2)	Prefers school to other activities OLS (3)	2002 test score OLS (4)
Program school * 2001 test score above winning threshold	0.13* (0.07)	0.09 (0.14)	-0.02 (0.07)	-0.08 (0.20)
Program school 2001 test score above winning threshold	0.03 (0.03)	-0.03 (0.07)	0.02 (0.02)	0.05 (0.12)
2001 test score polynomial controls	Yes	Yes	Yes	Yes
Age, school ethnic controls	Yes	Yes	Yes	Yes
Sample Size	1155	707	707	707
R ²	0.08	0.02	0.04	0.58
Mean of dependent variable	0.74	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 0.95 s.d.

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

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

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

Time	Activity
<u>2000</u>	
November	Fifth grade students in cohort 1 take district exams (these are baseline scores in the econometric analysis).
<u>2001</u>	
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).
<u>2002</u>	
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.
<u>2003</u>	
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)**

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

Notes: Significantly different than zero at 90% (*), 95% (**), 99% (***) confidence. 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 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, attendance, and test score data .