

Incentivizing Schooling for Learning

Evidence on the Impact of Alternative Targeting Approaches

Felipe Barrera-Osorio

Deon Filmer

The World Bank
Development Research Group
Human Development and Public Services Team
&
Education Team
East Asia and Pacific Region
July 2013



Abstract

This paper evaluates a primary school scholarship program in Cambodia with two different targeting mechanisms, one based on poverty level and the other on baseline test scores (“merit”). Both targeting mechanisms increased enrollment and attendance. However, only the merit-based targeting induced positive effects on test scores. The paper shows that the asymmetry of response is unlikely to have been driven by differences between

recipients’ characteristics. Higher student and family effort among beneficiaries of the merit-based scholarships suggest that the framing of the scholarship mattered for impact. The results suggest that in order to balance equity and efficiency, a two-step targeting approach might be preferable: first, target low-income individuals, and then, among them, target based on merit.

This paper is a product of the Human Development and Public Services Team, Development Research Group; and the Education Team, East Asia and Pacific Region. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at felipe_barrera-osorio@gse.harvard.edu or dfilmer@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Incentivizing schooling for learning: Evidence on the impact of alternative targeting approaches¹

Felipe Barrera-Osorio* and Deon Filmer**

JEL classification codes: I21; I24; I28; O10

Keywords: education; Cambodia; randomization; scholarships; merit-based targeting; poverty-based targeting.

* Harvard Graduate School of Education; ** World Bank

Sector Board: EDU

¹ We thank Luis Benveniste, Norbert Schady, Beng Simeth, Tsuyoshi Fukoaka, and the members of the Primary School Scholarship Team of the Royal Government of Cambodia's Ministry of Education for valuable input and assistance in carrying out this work. Adela Soliz provided able research assistance. The paper has also benefitted from comments by David Deming, Leigh Linden, Muna Meky, Richard Murnane, Halsey Rogers, Shwetlena Sabarwal, and Katja Vinha. The authors are responsible for any errors. This work benefited from funding from the World Bank as well as from the EPDF Trust Fund (TF095245). The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments they represent.

Introduction

Studies of the impacts of programs that incentivize schooling through cash transfers typically have found that direct and indirect costs are important determinants of school participation and that programs reducing such costs are quite effective in inducing higher enrollment and attendance rates (for a general review, see Fiszbein and Schady, 2009). Once students are enrolled in school, however, the programs do not show consistent positive impacts on learning outcomes.

One hypothesis that emerges from this literature (particularly the literature from developing countries) is that monetary incentives may increase the participation of low-achieving students in school, and that school systems may be ill-prepared to teach them well (Behrman, Parker, and Todd 2011; Filmer and Schady, 2009). This gives rise to a potential equity-efficiency tradeoff: Targeting high-achieving students through an incentive based on academic performance such as a merit scholarship may yield higher learning outcomes—but if academic performance is correlated with economic circumstance, then those outcomes would come at the cost of reaching the poor. Poverty-targeted incentives, on the other hand, do reach the poor and do induce greater schooling—but if there is little learning to show for it, one might question the usefulness of that approach. The setting for this study is schooling in Cambodia, a low-income country, yet the potential tradeoff between equity and efficiency has been at the center of discussions in many contexts: discussions on college scholarships in the United States (Orfield 2002), social programs in developing countries (Coady et al., 2003), and more generally on poverty reduction strategies in the presence of tight budget constraints (Bardhan, 1996).

This paper addresses the equity-efficiency tradeoff directly by evaluating the impacts of a scholarship program in Cambodia that included two approaches to targeting that were run in parallel. Schools in one group offered scholarships based on the economic status of the student's household (which we refer to as poverty-based targeting), and schools in another group offered scholarships based on the student's performance on a baseline test (which we refer to as merit-based targeting). Other than the mechanism for selecting the scholarship recipients, the conditions of the program under both targeting mechanisms were identical: Both had the same transfer amount and periodicity, both had the same delivery mechanism, and both had the same conditions for renewing the scholarship from one year to the next.

The random allocation of schools to each of these targeting mechanisms allows us to compare their relative impact on measures of school participation and achievement using a straightforward approach. Our evaluation shows that both targeting approaches increase enrollment and attendance rates, but only the merit-based scholarship shows a positive impact on achievement, measured by test scores. This

finding suggests that there is indeed a potential equity-efficiency tradeoff.² This tradeoff does not appear to be driven by the characteristics of the students themselves, however. High baseline achievers who received the poverty-based scholarship performed no better in follow-up tests than the corresponding control group. On the other hand, poor individuals who received a merit-based scholarship did perform better on the follow-up test.

This paper's findings are consistent with the notion that motivation is an important factor in a program's success in incentivizing students, and that the framing of the scholarship itself is a determining factor for motivation. Through framing, a particular treatment can label individuals a certain way, and the label can either reinforce or mitigate the impact of the treatment (Schmader, Johns, and Forbes 2008). By virtue of being labeled as academically successful, merit-based scholarship recipients may be motivated to work more, and their families to invest more; poverty-based recipients may not be motivated in a similar way. We find that, indeed, merit-based scholarship recipients and their families exert additional effort as a result of the scholarship—as measured by homework and by expenditures on education—whereas poverty-based scholarship recipients do not.

This paper is organized as follows. After discussing the pertinent literature in section 2, we describe the setting, the program, and the evaluation design in section 3. In section 4, we present the empirical strategy, the data, and the validation of the identification strategy. The main results appear in section 5. In section 6, we discuss the results, offer conclusions, and review the implications of the findings for designing future incentive programs.

1. Related literature: The efficiency margin of incentive programs

Monetary incentive programs (scholarships included) are thought to induce greater schooling for three main theoretical reasons. First, the direct and indirect costs of attending school, along with the lack of financing to cover those costs, may significantly deter families from making optimal decisions on education; reducing those costs may induce greater investment in education. Second, students and families may discount future returns to education very heavily, and as a result, they may not invest the optimal effort in education. Monetary incentives will increase the short-run benefits of such investments.

² This study narrowly defines efficiency as the effects of a program on educational outcomes per dollar cost of the program. Thus we do not investigate the effects on efficiency of raising the money for the program. The program transfers cash to poor households, which is potentially welfare-enhancing in itself, but we do not address this aspect of the program directly. Another issue that is not addressed or evaluated here is that academic achievement is but one objective—albeit an important one—of schooling, but schooling has additional social as well as personal impacts (for example, better health, delayed marriage).

Third, families and students may not have complete information on the returns to education. Monetary incentives may serve as a means of signaling that education is important.

Scholarship programs, such as the one evaluated here, can be thought of as a particular type of Conditional Cash Transfer (CCT) program—one that has an individual rather than a family as the recipient and that focuses on a single sector (education) rather than being conditional on actions related to two sectors (health and education).³ Extensively studied and increasingly popular in much of the developing world, CCT programs have become the largest form of social assistance provided in several countries (see the review in Fiszbein and Schady, 2009). Much of the rigorous evidence on the impact of CCTs has come from middle-income countries, mainly in Latin America, where baseline enrollment rates are high and impacts relatively small (in part because there is little room for an increase). In contrast, this paper presents evidence from a scholarship program in Cambodia, where enrollment rates are low. It adds to the much smaller evidence base on CCTs from low-income countries, such as Bangladesh (Chaudhury and Parajuli, 2008) and Malawi (Bair, McIntosh, and Ozler, 2009), as well as evidence from other programs in Cambodia (Filmer and Schady, 2008; Filmer and Schady, 2011).

CCT programs designed to increase enrollment and attendance have indeed raised these measures of school participation, but in some settings they have resulted in negative or insignificant changes in learning outcomes (Behrman, Parker, and Todd, 2005; Behrman, Sengupta, and Todd, 2000; Filmer and Schady, 2009), whereas in others they have had positive results (Baird, McIntosh and Ozler, 2011). Similarly, the effects of cash incentives linked to changes in learning itself have yielded mixed results. These programs aim to encourage the performance of students who are already in the educational system. Two recent evaluations in the United States showed mixed (Fryer, 2011) or no impacts (Bettinger, 2010) on standardized tests; an evaluation of a program in Israel showed positive impacts on exit-exams, take-up rates, and college matriculations (Angrist and Lavy, 2009). Finally, a mixed approach using merit-based scholarships in Kenya yielded positive effects on test scores in one of two districts (Kremer, Miguel, and Thornton, 2009).⁴ The question of how to effectively turn incentives for schooling into incentives for learning remains open.

³ Conditional Cash Transfer (CCT) programs transfer cash to families that comply with a set of conditions, such as the enrollment and regular attendance of children in school, regular prenatal visits by pregnant women, and regular health checkups for young children.

⁴ Besides giving scholarships concurrent with studies, some programs “promise” students a scholarship if they perform well on a test administered in the future. In theory, the promise of a reward elicits increased effort from students whose abilities place them within reach of a scholarship. See Kremer, Miguel, and Thornton (2009) for an evaluation of one such program.

In Cambodia, two evaluations of the impact of scholarships for lower secondary school have shown substantial increases in school enrollment and attendance as a direct consequence of the programs (Filmer and Schady, 2008 and 2011). Recipients were 20–30 percentage points more likely to be enrolled and attending school as a result of the scholarships. Filmer and Schady (2009) also show that scholarships targeted to lower secondary school students led to more expenditure on education and less work for pay among recipients. Impacts on learning outcomes were limited. The authors argue that the limited impacts point to potential issues in the quality of education and the match between students' skill levels (particularly among students induced to stay in school as a result of the scholarship) and instruction.

Three main explanations are typically put forward for mixed findings on test scores. As noted, the positive results may reflect the capacity of monetary incentives to act as an extrinsic motivator (for the student and family), especially among individuals from low-income families. Such short-term motivation may be important when information on the returns to education is imperfect and the discount rate is very high. Bettinger (2010) argues that these considerations may be highly relevant for primary school students in the United States. On the other hand, scholarship programs can induce a negative impact by reducing intrinsic motivation. A person's motivation to perform well can decline if she comes to view performing well as an obligatory activity for achieving a certain goal (in this case, a scholarship) (for example, see Lepper et al., 1973; Deci, Koestner, and Ryan, 2001). Finally, these programs may have no impact on academic achievement if students are unable to respond to the incentives. Students may not know how to convert the incentive into actions that influence achievement (Fryer, 2011). A related issue is that the incentive may have strong complementarities with other inputs that are out of a student's control and that the scholarship program does not alter, such as the quality or appropriateness of the teaching (Fryer, 2011): A student or her family may have a substantial amount of control over whether she attends school but far less control over the factors that enable that schooling to be converted into learning.

We explore an additional explanation that is consistent with the contrasting impacts of the poverty-based and merit-based targeting in Cambodia. If the targeting approach itself changes the frame within which the students are responding to the incentive, and that frame matters for impact, then the impact of the program on test scores will be dependent on the actual targeting mechanism. This line of reasoning is aligned with a (large) psychology literature on "stereotype threat" (for a general theoretical framework, see Schmader, Johns, and Forbers, 2008). When certain individuals are faced with a specific label that carries a social stigma, labeling interferes with intellectual performance. The mechanisms by which labeling can affect performance are diverse: anxiety, stereotype activation, self-doubt, working memory, and arousal (Schmader, Johns, and Forbes, 2008). For example, African Americans responded to stereotype threats by lowering performance in tests (Steele and Aronson, 1995). Labeling associated with

caste in India affected performance of lower-caste individuals when the caste was publicly announced (Hoff and Pandey, 2006). Aronson et al. (1999) shows that this effect can be present, even in the absence of a social stigma attached to the labeling; white males underperform in relation to Asian students.

If poverty-based scholarships trigger lower motivation and effort than a merit-based scholarship, not because of differences in the underlying skills of the population, but because of framing, this information will affect how the equity-efficiency tradeoff is made. Similarly, if the framing of merit-based scholarship triggers high motivation and self-esteem, then this targeting mechanism can induce efficiency gains.

2. Program and evaluation design

Cambodia has a recent record of using demand-side incentives to raise school enrollment and attendance rates. Some of these programs operate at the primary school level—such as school feeding programs or small-scale programs that offer incentives for children to attend school—but most are targeted at lower secondary school (Filmer and Schady, 2011). The programs do not simply waive school fees; rather, the families of children selected for a “scholarship” receive a small cash transfer, conditional on school enrollment, regular attendance, and satisfactory grades.

One important finding from the evaluations of earlier incentive programs is that their targeting only mildly favored the poor. Filmer and Schady (2009) show that one program, despite reaching the poorest children who *applied* for scholarships, did not reach the poorest of the poor, who had already dropped out of school before Grade 6—when they would apply for secondary school scholarships. Figure 1 presents the proportion of children ages 15–19 nationally who completed each grade at around the time the program evaluated in this paper was launched. The figure shows that children from the poorest quintiles are much less likely to reach 6th grade. This finding suggests that it is hard for a program that targets children at the end of Grade 6 to be strongly pro-poor—and that a program targeting poor students earlier in the schooling cycle is needed when the goal is to reach the poorest of the poor.

Primary School Scholarship Pilot Program

Based in part on these findings, and the desire to assess the viability, effectiveness, and optimal design of such a program, the Government of Cambodia began to implement a new pilot scholarship program in 2008. The program’s stated goal was to offset the direct and opportunity costs of education,

and increase schooling as a result.⁵ The implicit goal was also to improve learning outcomes through the additional schooling. This paper reports the results of the impact evaluation of that pilot program.

The basic design of the primary school scholarship pilot was to select participating schools; assign them randomly to offer scholarships based on poverty or merit; and then, within each school, identify candidates for scholarships based on transparent, clearly articulated criteria. Once selected, scholarship recipients were required to stay enrolled, attend school regularly, and maintain passing grades to keep the scholarship until they graduated from primary school.⁶ In Cambodia, primary school consists of Grades 1 through 6, and the program targeted students entering the upper-primary level (Grades 4, 5, and 6). The scholarship was equivalent to US\$ 20 per student annually.⁷ Through a follow-up survey, we determined that the mean household expenditure per capita per year in our sample was US\$ 610.24;⁸ as such, the scholarship represents 3.3% of the yearly per capita expenditure. The scholarships were intended to be disbursed in two tranches of US\$ 10—the first in the beginning of the year, and the second in the middle of the school year. In the first year of the program, however, scholarships were distributed in one lump sum due to delays in implementation.⁹

The pilot program was implemented in three provinces—Monduliri, Ratanakiri, and Preah Vihear—where average dropout rates between Grades 3 and 6 were highest, according to an analysis of Cambodia’s Education Management Information System (EMIS). To narrow the geographic scope of the program, only seven districts in Ratanakiri (those with the highest dropout rates) were selected for participation, out of a total of nine districts. In the other two provinces, all districts were included. Within these selected districts, all primary schools that offered classes through Grade 6 participated in the program.

The program offered scholarships to approximately 5,162 students from a pool of 12,078 individuals in 208 schools. Of the 208 schools, 103 were randomly assigned to join the program in Phase 1 (the

⁵ Primary schools are officially non-fee based. Opportunity costs include various forms of child labor which are relatively prevalent in the study area—although typically labor is combined with schooling at the primary school ages.

⁶ There requirements are moderately enforced. Students absent for many days are followed up by school officials and if they return to school they remain eligible for the scholarship. After a student is absent for too many days they are classified as having dropped out and no longer are eligible for the scholarship.

⁷ The previous lower-secondary scholarships were in the amounts of \$45 and \$60 per year—however, the evaluation found little impact on enrollment and attendance of \$60 over and above that of \$45 (Filmer and Schady, 2011)

⁸ GNI per capita in Cambodia was, approximately, UD\$700 for 2008 (World Development Indicators).

⁹ Scholarships distributions for the cohort of recipients analyzed here (Phase 1) took place in July 2009 (US\$ 20), November 2009 (US\$ 10), April 2010 (US\$ 10), November 2010 (US\$ 10); and April 2011 (US\$ 10).

program's first year, 2008/09) and the other 105 in Phase 2 (the second year of the program, 2009/10). In each phase, schools were randomly assigned to either poverty-based or merit-based scholarship targeting.

Schools using poverty-based targeting selected students based on a poverty index. All prospective recipients filled out a simple form with questions regarding their household and family socioeconomic characteristics.¹⁰ These forms were scored according to a strict formula based on weights derived from an analysis of household survey data to derive a poverty index for each student.¹¹ The poverty index ranged from 0 (richest household) to 292 (poorest household). The application forms were scored centrally by a firm contracted specifically for this purpose, thereby reducing the probability of any manipulation of program eligibility. Within each school, the applicants with the highest scores (that is, those with the highest level of poverty) were offered a scholarship.

Schools using merit-based targeting selected scholarship recipients based on scores on a test administered at baseline. The test was adapted from the Grade 3 National Learning Assessment.¹² All prospective recipients took the test, and within each school the applicants with the highest test scores were offered scholarships. Again, the tests were scored centrally to minimize the risk of program manipulation.

The number of students within each school who would receive a scholarship was fixed exogenously, and set to half the number of registered students in the year prior to the program (as determined by an analysis of EMIS data).¹³

Evaluation design and data

The identification of impacts relies on the fact that among the cohort of students studied, 4th graders in Phase 2 schools were not eligible for scholarships in the 2008/09 school year or later (Figure 2). Among those students, we identify the valid counterfactual group—a group that differs, on average, from

¹⁰ Table 1 reports the full set of variables included in the calculation of the score.

¹¹ The weights were determined by estimating a model predicting the probability that a student would drop out of school during Grades 4 to 6—since addressing this dropout was the stated goal of the program. Strictly speaking, the score should be referred to as a “dropout-risk score.” However, the risk is essentially a set of characteristics that capture the socioeconomic status of a household, weighted to capture those elements that predict dropout best. For convenience and ease of exposition, the score is referred to in this paper, as well as in program documents, as a “poverty” score.

¹² The National Assessment was implemented in a sample of schools nationwide in Grade 3 during the 2005/06 school year (Royal Government of Cambodia, 2006).

¹³ The number of scholarships is not equal to half the number of applicants for two reasons: First, the rule was to allocate scholarships to all applicants who had scores *higher than* or *tied with* the cutoff score; second, there were Grade 4, Phase 2 schools that did not receive the scholarship (control schools).

the treatment group only in that it did not receive scholarships. Since students in the control schools were never exposed to the program (even after the *subsequent* cohorts in Phase 2 schools became eligible for scholarships), the two groups of students can be tracked over time and their enrollment, attendance, and other outcomes compared.¹⁴ All prospective recipients in program schools filled out the application forms and took the assessment test. Recipients received scholarship disbursements (conditional on remaining in school, attending regularly, and maintaining passing grades) during the 2008/09, 2009/10, and 2010/11 school years.

We use three main data sources to evaluate program impacts. First, we use the full set of data collected at the time the students applied for the scholarships: data on baseline household characteristics (which were used to construct the poverty index), as well as mathematics and Khmer language test scores for all applicants. Second, we use the official list of students who were offered a scholarship. Third, we use endline data that were collected specifically for this evaluation. These data are derived from a survey administered at the end of the 2010/11 school year, three years after the program began implementation, to a random subsample of students from each program school. The students who began participation in Phase 1 and stayed in school were finishing (or had just finished) Grade 6 at the time of the survey. The survey was administered at home (not in schools) to the child who applied for the scholarship, and it included a household module administered to the child's mother, father, or other caregiver. In total 3,618 applicants were interviewed. For the majority of the analyses, we use data from 1,377 students from 204 schools¹⁵ who were in Grade 4 at baseline and were offered a poverty- or merit-based scholarship, or would have been offered a poverty- or merit-based scholarship had they attended a Phase 1 school.

The endline survey includes measures of school attainment, of the “intensity” of school participation (based on questions related to time spent in school), as well as two measures of achievement and cognitive development based on a mathematics and a “Digitspan” test. The items on the mathematics test were drawn from a variety of sources, including the baseline mathematics test, questions from the Grade 6 National Assessment, as well as publicly released items from the Trends in International Mathematics and Science Study (TIMSS) Grade 4 Assessment. A pretest ensured that only items with adequate properties were retained for the final test. The final multiple choice test measured both knowledge and the capacity to use this knowledge for problem solving. Performance on the math test is a measure of the most

¹⁴ Because the scholarship offers are made according to strict criteria, applicants “just above” and “just below” the cutoff for eligibility could be compared using a regression discontinuity design (RDD) approach to evaluate program impact. In future work we will use this approach, and we will be able to use data from the first and second cohorts of students who applied to the program.

¹⁵ Students from 4 of the original 208 schools in the sample were not part of the follow up survey because they did not offer Grade 4.

immediate academic impact of the intervention. The program is also expected to have an impact on endline test scores among students selected through each targeting mechanism for two main reasons. First, the program incentivized more enrollment and school attendance; consequently students are more exposed to school, and through that additional schooling they potentially acquire more learning. Second, because the program requires that all treated students—those with merit and those with poverty scholarships—maintain passing grades, the program is expected to incentivize students to study more, which in turn should affect their ability to solve mathematical problems.

In a Digitspan test, a series of numbers are read to a respondent who is then asked to repeat the numbers back to the enumerator. The series increase from 2 digits to a larger and larger number, up to 9-digit numbers. Respondents are also asked to repeat the numbers back in reverse order. The test is often included in batteries of psychometric tests, has been used in previous analyses of development programs (for example, see Kazianga, de Walque, and Alderman, 2012), and is typically interpreted as a measure of short term memory and working memory capacity.

Two particular characteristics of the data are noteworthy. First, we use follow-up information collected from home visits rather than at the school. This strategy allows us to avoid the problems that arise when follow-up information is collected at school, where some students may not be present. When follow-up information is collected at home, all students are included, not only those who attend school regularly. Second, we collect follow-up data three years after the students first receive the scholarship, which allows us to capture longer-run effects than most evaluations of the impact of school-related interventions typically manage to capture.

3. Empirical strategy

Empirical strategy

We estimate a reduced-form model relating the program to enrollment and attendance outcomes, test scores, and potential transmission mechanisms (school / teacher effort and student / household effort). A separate estimation is carried out for each targeting mechanism. For the poverty-based treatment, the estimation is based on the equation

$$Y_{i,t1} = \beta_0 + \beta_1 T_{i,s}^P + X'_{i,t0} B + \eta_{i,s,t1} \quad (1)$$

where $Y_{i,t1}$ denotes the value of the outcome variable (e.g., school attainment) for individual i at follow-up ($t1$); $X'_{i,t0}$ is a vector of control variables measured at baseline (e.g., household characteristics); and

$\eta_{i,t1}$ captures unobserved student characteristics and idiosyncratic shocks. $T_{i,s}^P$ is equal to one if the student was offered the poverty-based scholarship, and zero for students in the control schools. The sample consists only of recipients in treatment schools, and individuals who would have been recipients of poverty-based scholarships had they attended a treatment school:

$$T_{i,s}^P = \begin{cases} 1 & \text{if } Poverty\ Index_i \geq Poverty\ Index_{50\ Perc, Poverty\ School} \\ 0 & \text{if } Poverty\ Index_i \geq Poverty\ Index_{50\ Perc, Control\ School} \end{cases}$$

An analogous estimation is done for the merit-based treatment:

$$Y_{i,t1} = \beta_0 + \beta_1 T_{i,s}^M + X'_{i,t0} B + \eta_{i,s,t1} \quad (2)$$

where T_i^M indicates treatment status for student i in merit-based school s . It is equal to one if the student was offered the merit-based scholarship, and zero for the sample of students in the control schools who would have been recipients of merit-based scholarships had they attended a treatment school:

$$T_{i,s}^M = \begin{cases} 1 & \text{if } Test\ Score_i \geq Test\ Score_{50\ Perc, Merit\ School,} \\ 0 & \text{if } Test\ Score_i \geq Test\ Score_{50\ Perc, Control\ School,} \end{cases}$$

The control variables ($X'_{i,t0}$) included in the estimation consist of an indicator for gender; the number of minors in the household; indicators for whether the household owns a motorcycle, a car/truck, an ox/buffalo, a pig, an ox or buffalo cart; indicators for whether the house has a hard roof, a hard wall, a hard floor, an automatic toilet, a pit toilet, electricity, piped water; and the poverty index and test scores.¹⁶ Note that since we are relying on randomized assignment for identification, the inclusion of control variables should not affect the estimate of impacts. We include them because they may potentially increase the precision of the estimates.

Given that the treatment variable identifies at the baseline all individuals who were offered the scholarship, the estimator is an intention-to-treat estimator (ITT). Errors are clustered at school level, and each model includes district-level fixed effects.

The design of the intervention also allows us to directly estimate the effects of the program on nontreated students (spillover effects from the treatment). In each treatment school, approximately half of the students did not receive scholarships. Given the random assignment of schools into treatment and

¹⁶ We use dummy indicative variables to impute values for a very small number of missing observations in the control variables. Given the random nature of the program, the imputation does not make any substantive difference in the estimation, although standard errors are slightly smaller since sample size is maximized.

control groups, we can compare the nonpoor (and nontreated) students in the poverty treatment schools with nonpoor (and nontreated) students in the control schools. Similarly, we can compare students who did not receive scholarships in the merit-based schools with students in control schools who would not have qualified for a merit scholarship. To estimate the effects on the nontreated students, we estimate models similar to those in Equations (1) and (2), but adjust the definition of T_i and the sample used in the analysis to fit this alternative specification. Since the scholarships were offered after the baseline test, and nonrecipients could not change their status during the duration of the program, any effects on nontreated students emanate from complementarities and interactions between treated and nontreated students in the treatment schools during the academic year.¹⁷

Baseline balance and characterization of the study sample

This section presents the general characteristics of the study sample and describes the validation of the experimental design. The first column in Table 1 reports the mean characteristics of applicants and their households in control schools (with standard deviations in parentheses). For example, Column 1 indicates that half the applicants were girls, applicants have on average 1.69 young siblings, few applicants come from households with a car (16 percent), and about half come from households that own an oxen or buffalo (55 percent). Column 2 reports the estimate of the difference between the characteristics of these applicants and the characteristics of all applicants in treatment schools, based on estimating:

$$X_{i,t10} = \gamma_0 + \gamma_1 T_{i,s} + \zeta_{i,s,t0} \quad (3)$$

on the follow-up sample at baseline (with the estimate of γ_1 reported in Table 1). For none of the indicators is there a statistically significant difference between the two sets of applicants, suggesting that the randomized allocation of treatment status to schools was successful at creating a valid experimental design.

Columns 3–6 narrow the sample for this analysis in two ways. First, the analysis is done separately for schools using poverty-based and merit-based targeting. Second, it includes only treated students in treatment schools and untreated students in control schools who would have been eligible for treatment, based on their poverty index score or baseline test score, had they attended a treatment school. The

¹⁷ These peer effects are different in nature from the externality effects estimated in Kremer et al. (2009). There the authors estimate the effect of the “promise” of a scholarship for students with low scores prior to the intervention. Those effects presumably emanate from the effort that all students may exercise to get the scholarship and also from the pressure that parents may exert on teachers.

sample is restricted to Grade 4 students at baseline, and as described, the control group of students were not beneficiaries of the program's expansion. Columns 3 and 5 report the means (and standard deviations) of household and individual characteristics prior to the intervention for each of the control groups, and Columns 4 and 6 report the estimated difference (with standard errors in parenthesis) between treatment and control applicants.

Overall, characteristics differ very little between treatment and control groups. Only a few coefficients in Columns 4 and 6 are statistically significant: Of the 32 differences reported, 5 are significant at the 10% level; among them, 1 is significant at the 1% level, a result that is consistent with pure chance. Again, the results from these tests confirm the validity of the random assignment, since both control and treatment groups are similar in their observed characteristics.

An important feature of the data is that the poverty index and the test score are statistically indistinguishable between the treatment and control groups. Figures 3 and 4 present the density of poverty index and the test score at baseline for treatment and control schools. There is a clear overlap throughout the whole distribution. A Kolmogorov-Smirnov test does not allow us to reject equality of both distributions.

Table 1 also shows that the applicants who are offered scholarships on the basis of merit have more assets, a lower poverty index score, and performed better on the baseline test than students offered a poverty-based scholarship. For instance, the poverty index, which ranges from 0 (wealthiest family in the sample) to 292 (poorest family in the sample), has a mean of 245.13 for students targeted based on poverty and 218.2 for students targeted based on merit. Likewise, the baseline test (ranging from 0 to 25) has a mean of 19.77 for merit-based students and 17.74 for poverty-based students. At the same time, however, the correlation coefficient between the poverty index and the test score for the entire sample of individuals at baseline is remarkably low (for example, the Spearman correlation coefficient is 0.16, and the pairwise correlation is 0.17). We return to this issue when we discuss the tradeoffs between the two targeting approaches. The main finding from Table 1, however, is that the samples are balanced at baseline, allowing us to identify causal effects.

Attrition

Before turning to the estimates of program impact, it is useful to discuss the construction of our sample. Recall that the endline analysis is based on a household survey administered to a random subsample of applicants. The survey company received a list of 4,225 randomly selected applicants to

visit at follow-up, as well as a list of potential (randomly generated) “replacement” students to be interviewed if the original person selected could not be found. The survey company located 3,191 students on the original list and 414 replacement students, yielding a total of 3,605 observations at follow-up and a 24.5% (1,034 out of 4,225) rate of attrition. Table 2 compares the characteristics of the 1,034 attritors (4,225 – 3,191) and the rest of the applicants at baseline. Columns 1 and 2 report the means (and standard deviations in parentheses) of household baseline characteristics for the baseline sample of nonattrited and attrited applicants; Column 3 reports the estimate (standard errors in parentheses) and statistical significance of the estimate of the difference between the two groups. In Columns 4 and 5, we test for differences between attritors of the treatment group and attritors of the control group (a difference in difference model) for each experiment.

Attritors differ from the full baseline sample in that attritors appear to be poorer, but the level of attrition did not vary in the treatment and control groups in either experiment (Columns 4 and 5). In short, the two research groups (poverty-based and merit-based) are balanced.

Column 6 in Table 2 presents the means of the household characteristics for the 3,191 students on the original list, and Column 7 shows the respective means for the 414 replacement students. Column 8 reports the differences between the two groups. In no case is there a statistically significant difference between the two groups.

We take these results to suggest that although the rate of attrition from the sample was relatively high (although not particularly higher than in similar studies), and although some differences exist between applicants who were and were not found, those differences were not systematically different in the two experiments in the study. Consequently we do not expect results on the differential impacts of these targeting approaches to be biased.

4. Results

Impacts on enrollment and attendance

The intervention is aimed at directly incentivizing higher enrollment and attendance rates, so scholarship students must stay enrolled, attend school regularly, and maintain passing grades until they graduate from primary school (6th grade). We focus on three enrollment and attendance proxies: the proportion of students reaching 6th grade, the highest grade completed, and the hours of school attended in the past seven days.

Table 3 reports the program’s impact on these enrollment and attendance proxies. Panel A presents the results for the poverty-based intervention—coefficient β_1 , Equation (1)—and Panel B presents the analogous estimates for the merit-based intervention—coefficient β_1 , Equation (2). For each panel, the first column presents the results without controlling for baseline characteristics, and the constant of the regression is the mean of the control group. For each panel, the second column presents the effects of the impacts after controlling for baseline student characteristics (i.e., the variables reported in Table 1), poverty index, test scores at baseline, and district fixed effects. As mentioned, the randomized assignment was successful in that it produced a balanced sample, and as expected, the results are very similar before and after controlling for baseline characteristics.

The mean of the control group for each outcome variable is a reference point for assessing the magnitude of the impacts. Just over 60% of the control students reported reaching at least 6th grade, and the average grade completion was around 5.4 grades. The third outcome variable was based on the number of hours students had attended school during the past seven days, conditional on being enrolled. On average, students reported having attended school for about 8.83 hours (the poverty-targeted group) and 9.27 hours (the merit-targeted group) in the past week.¹⁸

Impacts on all enrollment and attendance variables were significant and positive in the poverty-targeted treatment. Poverty-based scholarship recipients are about 18 percentage points more likely to reach Grade 6 than control students; they complete just over 0.33 more grades, and they attended school around 3 hours more than the control group. Similar effects are found in the merit-based intervention: Students in the merit-targeted treatment group are 13 percentage points more likely to reach Grade 6 than control students; they complete about 0.2 more grades than controls; and they attend about 1 additional hour of school compared to the control students (however, this last coefficient is not statistically significant). These impacts are comparable to those found in the context of the Lower Secondary School scholarship program, where enrollment increased on the order of 20–25 percentage points (Filmer and Schady, 2011). These impacts are larger than the majority of impacts documented in countries elsewhere in the world (Fiszbein and Schady, 2011), and they are also large when assessed against the very small size of the transfer (US\$ 20 per year).

In sum, there is strong evidence that the program increased school participation regardless of whether the scholarship was awarded based on merit or poverty.

¹⁸ In cases where the school was closed because classes had ended for the year, we assigned a value of zero, because we did not know exactly when the school year had ended. We do not expect that school closures were systematically different in any of the groups studied.

Results on test scores

Table 4 presents impacts of the interventions on the mathematics test and the Digitspan test, the two measures of academic and cognitive achievement. Table 4 has the same structure as Table 3: It presents impact estimates from models without controls and after controlling for baseline characteristics—including baseline test score and poverty index—and district fixed effects. Panel A presents the effects of the poverty-based treatment, and Panel B presents those for the merit-based treatment. All achievement measures are standardized such that the control schools have a mean of zero and a standard deviation of one (when averaged across all students, not just those who serve as the counterfactual for the treatment group). Impacts can therefore be interpreted as changes in a standard deviation of the achievement measure.

In contrast to the results on enrollment and attendance, the impacts on test scores differ between recipients of poverty-based and merit-based scholarships (Table 4). Poverty-based scholarships had no impact on test scores, whereas merit-based scholarships had positive impacts. For the models with controls, the impacts of the poverty-based scholarships on the mathematics and Digitspan test scores are estimated to be very small (on the order of half a standard deviation) and negative, and they are not statistically significantly different from zero. On the other hand, the estimates of the merit-based scholarships are 0.170 standard deviations for the mathematics test and 0.149 standard deviations for the Digitspan test. As discussed, the impacts on mathematics are likely to be a direct effect of greater exposure to schooling and perhaps increased effort. Note that these test score impacts are similar in magnitude to those reported for the merit-based scholarship program evaluated in Busia, Kenya (Kremer, Miguel, and Thornton, 2009). In contrast, results for the Digitspan test are less likely to be directly related to learning in school and potentially capture the effect of labeling on working memory. Indeed, Schmader, Johns, and Forber (2008) argue that labeling can affect working memory. In this case, our results are consistent with merit-based scholarships producing a positive effect on working memory.

These results suggest that both merit-based and poverty-based scholarships gave students an incentive to acquire more schooling. Only students who received merit-based scholarships showed any gains in academic achievement, however.

As discussed, the different results for recipients of poverty-based and merit-based scholarships could potentially derive from the different skills and endowments of individuals in these groups at baseline. We can test this hypothesis directly, because some recipients of poverty-based scholarships performed well enough on the baseline test to qualify for merit-based scholarships, had they been offered. These high

performers in the poverty-based treatment are identical to the high performers in the merit-based treatment in every respect but one: Their scholarships were labeled “poverty” scholarships rather than “merit” scholarships. Absent any labeling effects, one would expect to find the same impacts on learning outcomes among these high achievers as among recipients of merit-based scholarships who are similarly poor. This is not what we find.

For each school where scholarships were offered based on poverty status, we identify students that were above and below each school’s median baseline test score and estimate separate regressions for these two samples. Similarly, for schools that awarded scholarships based on merit, we identify students above and below each school’s median poverty index and estimate separate regressions for each sample.¹⁹ Table 5 presents the estimation results for poverty-based scholarships (Panel A) and merit-based scholarships (Panel B).

Impacts for the poverty-based scholarships, across all groups, are not statistically significantly different from zero. The impact of these scholarships on mathematics test scores among the low and high baseline achievers are, as in the overall sample, small and negative (-0.019 and -0.069, respectively). In contrast, the merit-based scholarships have a small but positive impact for the nonpoor (impact of 0.053 standard deviations), although this estimate is not statistically significant; and a large positive impact for the poor (impact of 0.233 standard deviations).

The results for the Digitspan test are of particular interest, since they reinforce the notion that labeling plays an important part in generating these results. If all of the effects seen here were driven only by greater exposure to school or greater academic effort, we would expect to find impacts primarily on mathematics test scores. If framing is important, it should have an impact on working memory, as put forward by Schmader, Johns, and Forber (2008). The results show small, negative, and statistically insignificant impacts of poverty-based scholarships on the Digitspan test: The impact among the low and high baseline achievers in the poverty-based treatment is -0.077 standard deviations for both groups. In contrast, the impact for the merit-based treatment is positive for both groups, nonpoor and poor, and similar in magnitude (0.194 standard deviations for the nonpoor and 0.142 standard deviations for the poor—although only the former is statistically significantly different from zero).

In short, the poverty-based scholarships did not induce better test results among recipients who also had high test scores at baseline, whereas the merit-based scholarships did induce better test scores among

¹⁹ The results are corroborated with an interaction model, in which the outcome variable is run against the treatment effect, a dummy that identifies the two separate samples, and the interaction term.

individuals who were poor at baseline. This finding is consistent with the framing of the scholarships (and the associated labeling) acting as important determinants of impacts in the program.

Effort

It is important to recognize that the scholarship program could influence the behavior of the school and teachers. For instance, under an altruistic model, teachers can give more attention to students with scholarships to help them retain their scholarships. It is also possible that parents of scholarship recipients may pressure the children's teachers to exert more effort. Banerjee and Duflo (2006) discuss changes in teacher motivation arising from greater accountability to families in the context of a scholarship program. In addition, teachers might put more effort into students who are labeled high performers (so-called "Pygmalion effects"; see Jacobson and Rosenthal 1968). As such, school actors (such as teachers) may change their behavior differentially with the introduction of different kinds of scholarships.

Although our data do not allow an analysis of these supply-side effects,²⁰ both types of scholarships studied here plausibly provide incentives for students to increase effort. We use two measures, based on the data available from household interviews, to assess the program's impact on effort through channels external to the school: the hours that students study outside of school, and the share of family expenditures going to education-related items such as textbooks.

The estimates of impacts on these two variables are presented in Table 6. As a reference point, students in the control group spent around 3.4 hours per week doing school tasks outside of school, and total household expenditures on education averaged US\$ 20. Merit-based scholarship recipients spent more time doing academic work outside of school compared to the control group (an increment of 0.7 hours); the impact for poverty-based scholarship recipients is almost half the size (0.4 hours) and statistically insignificant. Similarly, spending on education differs by type of scholarship received. Households with a merit-based scholarship recipient spent almost US\$ 6 more on education than control students' households, whereas households that received a poverty-based scholarship did not discernably increase expenditures on education.

In sum, it appears that students work harder, and families spend more on education, when a merit-based scholarship is awarded. No such impacts are seen among recipients of poverty-based scholarships.

Peer effects on enrollment, attendance, and achievement

²⁰ The current consensus of the literature on the Pygmalion effect is that it exists but is small. Teachers also tend to correct information as the academic year progress (Jussim and Harber, 2005)

As discussed, the design of the intervention allows us to test directly for peer effects. Because the scope for containing spillover effects is always limited, they should be estimated directly and factored into any assessment of the effectiveness of a program such as this one.²¹

The Cambodian scholarship program evaluated here has several potential sources of peer effects. If increased enrollment and attendance lead to overcrowded classrooms, the learning opportunities afforded to other students may be harmed. At the same time, spillover effects might be positive if scholarships create an environment that favors schooling and learning among and affects all children in a classroom.²² Peer effects might differ depending on the targeting approach. For instance, applicants denied a merit-based scholarship might become discouraged and eventually perform worse. If teachers respond differently to students based on the type of scholarship they receive, the estimation of peer effects could potentially capture part of that response (to the extent that it manifests in outcomes among peers). For example, it may be possible that in the merit-based schools, teachers focus instruction and pedagogy on scholarship recipients to the detriment of nonrecipients. Similarly, teachers in schools where scholarships were awarded based on poverty levels could potentially dedicate more effort to their nonpoor students. These effects would show up in the estimation of peer effects.

Table 7 presents coefficient estimates where the treatment indicator variable is set at one for students who received no scholarships in schools where poverty- or merit-based scholarships were awarded, and zero for students in control schools who had analogous baseline test scores or poverty index scores (as before, the estimation sample consists only of these two sets of applicants). The dependent variables are the same attendance and achievement variables used before: The proportion of students reaching 6th grade, the highest grade completed, and the hours of school attended. The results suggest the presence of very limited peer effects. The only statistically significant coefficient is in the probability of reaching Grade 6, which is 8.2 percentage points higher in schools where scholarships were offered based on poverty. All of the other peer-effect impacts (highest grade completed and number of hours in school for the poverty-based scholarships, and all outcomes for the merit-based scholarships) are estimated to be small and not statistically significantly different from zero. The results dispel the notion that merit scholarship applicants who did not receive a scholarship became discouraged and performed worse than they would have otherwise performed.

²¹ As discussed in footnote 17, these peer effects are different in nature from the externality effects estimated in Kremer et al. (2009).

²² An additional positive spillover might occur among younger cohorts who stay in school longer with the hope of eventually benefiting from a scholarship. We have no data to address this issue, however.

Table 8 is the analogous table for impacts of the mathematics and Digitspan tests among nontreated peers. None of the coefficient estimates are statistically significant, and all the coefficient estimates are very close to zero, suggesting that no negative—or positive—spillovers are associated with either targeting approach.

In sum, it seems that poverty-based scholarships may induce more school attendance among nonrecipients. The program appeared to have no other peer effects, positive or negative.

Effects on equity

Given the two targeting approaches, the Cambodia pilot scholarship program is well suited to shed light on the potential tradeoff between efficiency—defined as achieving more per dollar transferred—and equity—defined as reaching the poorest population. Analyses of the socioeconomic profile of program applicants and recipients under the two targeting schemes, and comparisons to the national distribution of socioeconomic characteristics, suggest that both targeting approaches are heavily skewed to the poor. The first panel in Figure 5 shows that 50% of those who applied to the program are in the poorest nationally-benchmarked quintile (70% in the poorest two quintiles); fewer than 3% of applicants were from the richest quintile. Clearly the program was targeted to poor areas and poor schools in Cambodia. Unsurprisingly, scholarships targeted to the poorest from each school yields a greater pro-poor distribution of benefits than when targeting by merit. In the poverty-targeted schools, 63% of applicants were from the poorest quintile of the population (85% were in the poorest two quintiles). Merit-based targeting is less pro-poor but is still able to reach the poorest groups in the population. In the merit-targeting schools 54% of applicants were from the poorest quintile of the population (and 76% in the poorest two quintiles).

The finding that merit-based targeting did not result in an overall regressive scheme is a reassuring result. Figure 6 shows that within schools, the association between poverty and test scores is likewise not as close as might have been feared. In this figure, the horizontal axis is the relative poverty ranking of an applicant, where 0 is the 50th percentile, +1 is the applicant ranked one position higher on the poverty scale, and -1 is the applicant ranked one position lower, and so forth. The vertical axis is the analogous relative rankings on the merit scale.

If only wealthier children were to score high on the merit test, and poorer children low, then all the observations would be in quadrants (A) and (D) of Figure 6. Clearly this is not the case: The

observations are roughly equally distributed across the four quadrants.²³ This means that a merit-based approach (which targets children in quadrants A and B) includes children from both wealthier backgrounds (quadrant A) as well as children from poorer backgrounds (quadrant B). Similarly, a poverty-based approach includes both higher-scoring (quadrant B) and lower-scoring (quadrant D) applicants.

These school-specific rankings are consistent with the benefit incidence analysis. The equity/efficiency tradeoff between poverty- and merit-based targeting is not particularly stark. Given the relatively effective geographic targeting of the program, it is unclear whether this result is generalizable. In other settings—for example, where there is greater heterogeneity in student poverty levels—the result may not hold. To the extent that the result is somewhat generalizable, the findings suggest that rather than there being an equity/efficiency tradeoff, there is perhaps a way to enhance both equity and efficiency.

5. Conclusions

The fact that some students were able to take better academic advantage than others from additional school exposure highlights an issue rarely addressed in previous evaluations of CCT programs. Recent evidence on monetary incentives for schooling shows that students are able to change their behavior on the margins that are under their control—for example, enrollment and attendance. These positive effects do not necessarily translate into test score gains, however. For example, despite the fact that Mexico's Oportunidades program—a rigorously evaluated CCT program—induced students to enroll and attend additional years of school, the program did not induce better test scores. A recent set of papers has argued that education systems in developing countries are typically tailored towards better-off and better-skilled students. Specifically, Glewwe, Kremer, and Moulin (2009) show that only the strongest students at baseline were able to take advantage of textbooks that were provided to schools in Kenya. Furthermore, Duflo, Dupas, and Kremer (2011) show that teachers who were assigned to students at the bottom of the achievement distribution were less likely to teach.

The most important finding of this study is the asymmetry of response of the two targeting mechanisms. Both poverty-based and merit-based targeting schemes induce higher enrollment and attendance, yet only the merit-based mechanism induces positive effects on test scores, and the merit-based students (and their households) exhibit a higher level of effort in education. These results are not driven by differences in baseline skills and preparation across the two types of scholarship students.

²³ In fact the regression line for this figure has a mildly *positive* slope: The regression of relative merit ranking versus relative poverty ranking yields a coefficient of 1.2 with a standard error of 0.17 (significant at the 1% level).

Poorer students who are not academically prepared are unable to gain significantly in learning (measured by test scores) from the additional schooling they receive as a result of the simple poverty-targeted incentive. Clearly more work is needed—in Cambodia and elsewhere—to establish the best approach to ensure that additional schooling translates into learning for these students. Remedial lessons for students in the early grades, or increasing school readiness among poorer students, for example through early child development programs, might be approaches to try. Indeed, data from Cambodia suggest that children suffer from substantial delays in cognitive development, which hamper their school readiness (Naudeau et al., 2011).

The experience evaluated in this paper suggests another approach, based on the finding that incentivizing school attendance in a way that recognizes academic potential can pay off in measurable learning outcomes. By changing the framing of the program, and the associated labeling of the recipient, the merit-based scholarships distributed by the program appear to have motivated recipients and their families to exert greater effort in education, with measureable impacts on learning outcomes.

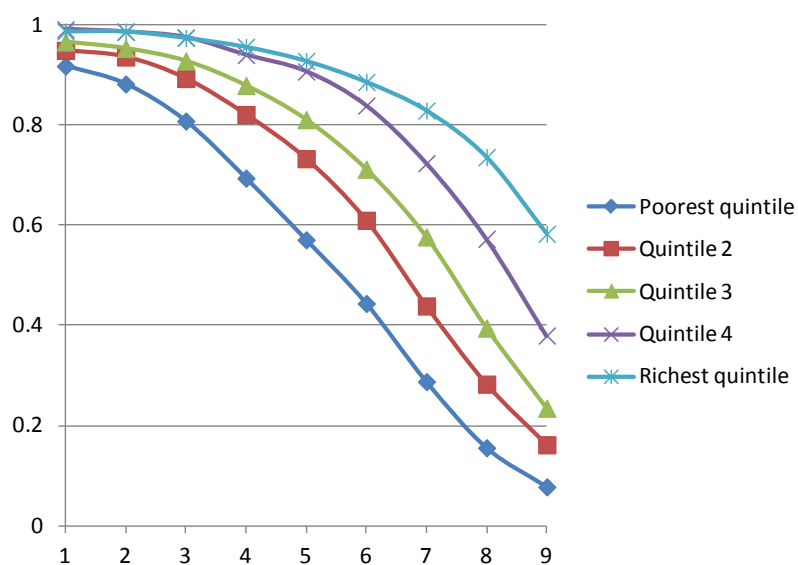
While the findings cannot answer all potential questions—for example, does teacher behavior also respond to the framing of the scholarships?—the results suggest that it should be possible to frame demand-side incentive programs in a way that maximizes their impact on learning outcomes. Such an approach could be undertaken in a way that does not necessarily imply a tradeoff between equity and efficiency. Indeed, the results show that among poorer children who received merit-based scholarships, the impacts on school participation as well as test scores were large. Scaling up an approach that targets students with an incentive that recognizes their high academic potential—while ensuring that the poorest students are among that set—is likely to maximize both the equity and effectiveness objectives of the program.

References

- Angrist J. and V. Lavy. 2009. "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial" *American Economic Review*, 99:4, pp. 1384-1414
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. 1999. "When white men can't do math: Necessary and sufficient factors in stereotype threat." *Journal of experimental social psychology*, 35(1), 29-46.
- Baird, Sarah, Craig McIntosh and Berk Özler, 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment," *The Quarterly Journal of Economics*, 126(4), pages 1709-1753.
- Bardhan, P. 1996. "Efficiency, equity and poverty alleviation: policy issues in less developed countries." *The Economic Journal*, 106(438), pages 1344-1356.
- Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives*, 20(1): 117-132.
- Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2005. "Long-Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." Discussion Paper 122, Ibero-America Institute for Economic Research, Göttingen, Germany.
- Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2011 "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/Oportunidades." *Journal of Human Resources*, 46, no. 1: 93-122.
- Behrman, Jere R., Piyali Sengupta, and Petra Todd. 2000. "The Impact of PROGRESA on Achievement Test Scores in the First Year." Unpublished manuscript, International Food Policy Research Institute, Washington, DC.
- Bettinger, Eric. 2010 "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores," NBER Working Paper 16333.
- Chaudhury, Nazmul and Dilip Parajuli. 2008. "Conditional Cash Transfers and Female Schooling: The Impact of the Female School Stipend Programme on Public School Enrolments in Punjab, Pakistan." *Applied Economics*, 42(28): 3565-3583.
- Coady, D., Grosh, M., and Hoddinott, J. 2002. *The targeting of transfers in developing countries: Review of experience and lessons*. Social Safety Net Primer Series, World Bank, Washington DC.
- Deci, E. L., Koestner, R., and Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, 71(1), 1-27.
- Duflo, Esther, Pascaline Dupas and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*. 101(5): 1739-74.
- Filmer, Deon, and Norbert Schady. 2008. "Getting Girls into School: Evidence from a Scholarship Program in Cambodia." *Economic Development and Cultural Change*, 56(2): 581-617
- Filmer, Deon and Norbert Schady. 2009. "School Enrollment, Selection and Test Scores." World Bank Policy Research Working Paper No. 4998. The World Bank, Washington, DC.

- Filmer, Deon and Norbert Schady. 2011. "Does more cash in conditional cash transfer programs always lead to larger impacts on school attendance?" *Journal of Development Economics*. 96(1): 150–157.
- Fiszbein, Ariel and Norbert Schady. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. The World Bank. Washington, DC
- Fryer Jr., Roland G. 2011. "Financial Incentives and student Achievement: Evidence from randomized trials." *The Quarterly Journal of Economics* 126, 1755–1798.
- Glewwe, Paul, Michael Kremer and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics*. 1(1): 112-135.
- Hoff, K. and P. Pandey. 2006. "Discrimination, Social Identity, and Durable Inequalities" *American Economic Review*, Vol 96, No. 2, pp 206-211
- Jussim, L., & Harber, K. D. (2005). "Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies." *Personality and Social Psychology Review*, 9(2), 131-155.
- Kazianga, Harounan, Damien de Walque and Harold Alderman. 2012. "Educational and Child Labour Impacts of Two Food-for-Education Schemes: Evidence from a Randomised Trial in Rural Burkina Faso." *Journal of African Economies*. 21(5): 723-760.
- Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics*. 91(3): 437-456.
- Lepper, M. R., D. Greene, and R. Nisbett. 1973. "Undermining Children's Intrinsic Interest with Extrinsic Rewards: A Test of the "Overjustification" Hypothesis" *Journal of Personality and Social Psychology*, Vol. 28, No. 1, pp. 129-137
- Naudeau, Sophie, Sebastian Martinez, Patrick Premand, and Deon Filmer. 2011. "Cognitive Development among Young Children in Low-Income Countries" in Alderman, Harold ed. *No Small Matter: The Impact of Poverty, Shocks, and Human Capital Investments in Early Childhood Development*. The World Bank. Washington, DC.
- Orfield, Gary, "Foreword," in Donald E. Heller and Patricia Marin (Eds.), *Who Should We Help? The Negative Social Consequences of Merit Aid Scholarships* (2002) (Papers presented at the conference "State Merit Aid Programs: College Access and Equity" at Harvard University). Document found in <http://civilrightsproject.ucla.edu/research/college-access/financing/who-should-we-help-the-negative-social-consequences-of-merit-scholarships/>
- Royal Government of Cambodia. 2006. Student Achievement and Education Policy: Results from the Grade Three Assessment—Final Report. Cambodia Education Sector Support Project—National Assessment Component. Phnom Penh, Cambodia.
- Schmader, T., Johns, M., and Forbes, C. (2008). "An integrated process model of stereotype threat effects on performance." *Psychological Review*, 115(2), 336.

Figure 1: Proportion of 15 to 19 year olds who have completed each grade, by quintile.



Source: Authors' analysis of data from Cambodia Demographic and Health Survey 2010.

Figure 2. Evaluation design

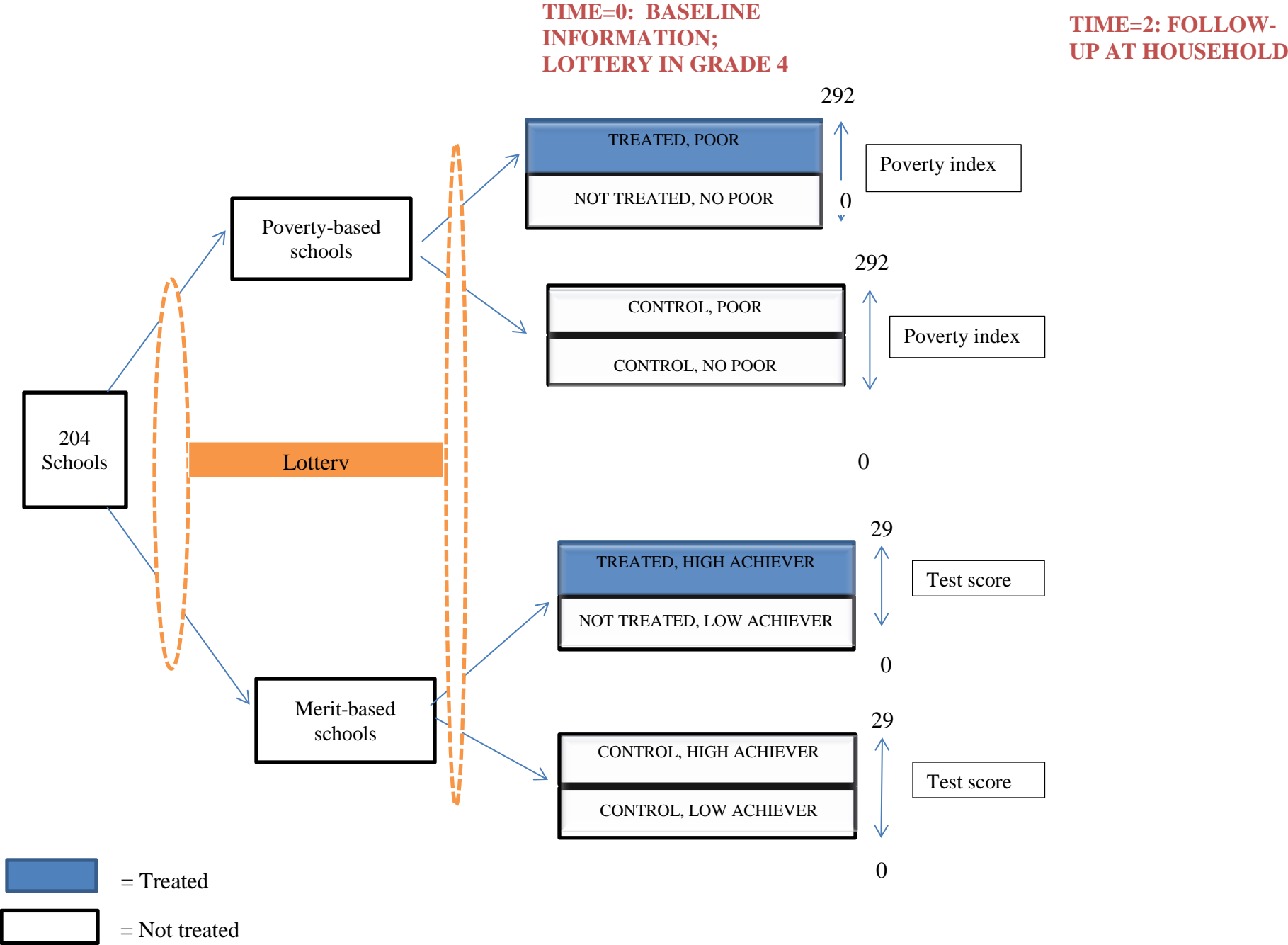
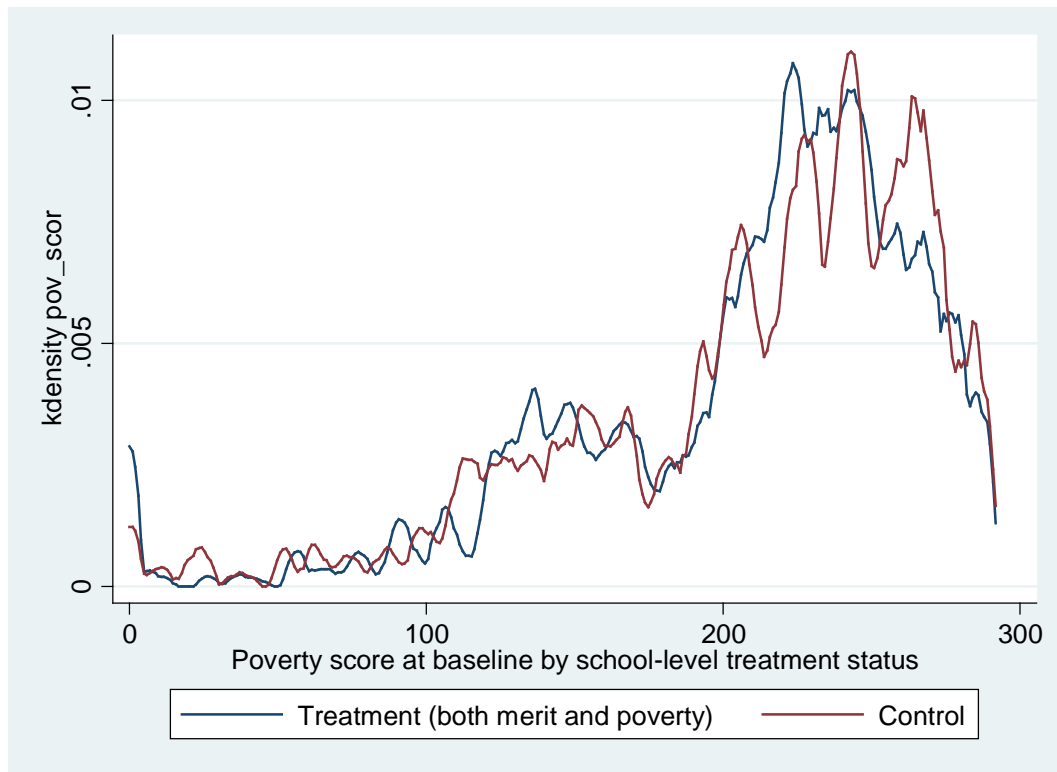
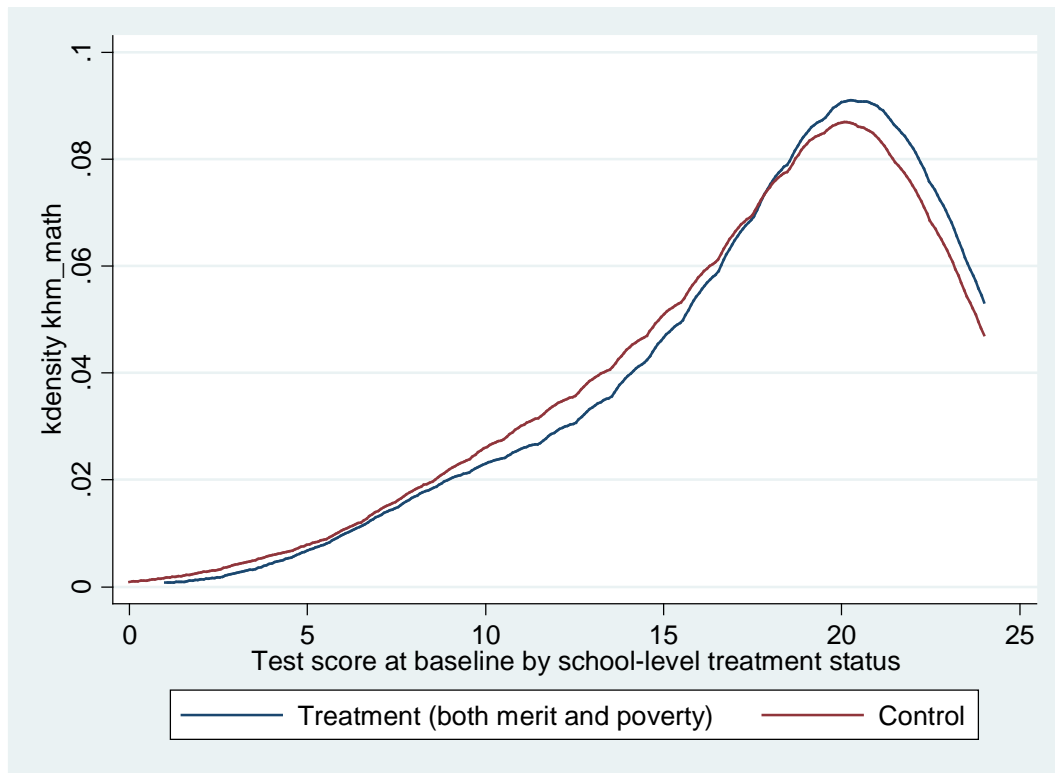


Figure 3: Poverty score at baseline, Treatment versus Control



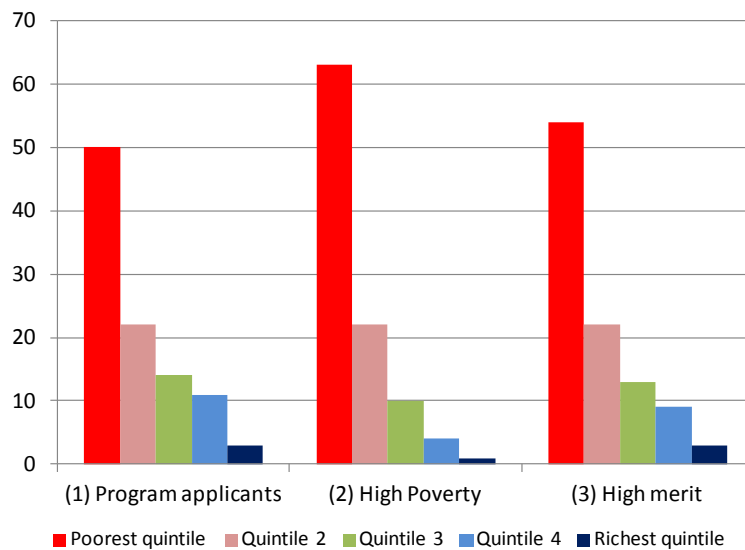
Source: Analysis of baseline application forms.

Figure 4: Test scores at baseline, Treatment versus Control



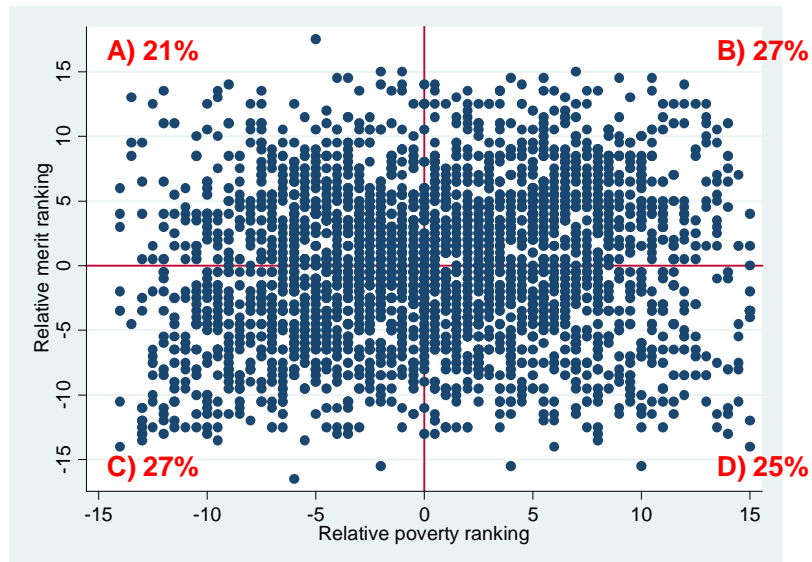
Source: Analysis of baseline Math and Khmer language tests.

Figure 5: Distribution of selected populations across nationally benchmarked quintiles



Source: Analysis of Cambodia DHS 2010 and Primary Scholarship Application forms. Quintiles are defined on the basis of an index of household wealth-related variables that are collected in both the DHS 2010 as well as on the scholarship program application forms.

Figure 6: The association between applicants' relative poverty and relative merit rankings



Source: Analysis of baseline application forms and baseline Math and Khmer language tests.

Table 1. Baseline Balance and mean and standard deviation of baseline characteristics

	School Level		Student Level			
	Control	Difference with treatment	Control Poverty	Difference with treatment	Control Merit	Difference with treatment
	(1)	(2)	(3)	(4)	(5)	(6)
Girl	0.49 (0.50)	0.037 (0.02)	0.52 (0.50)	0.100*** (0.03)	0.49 (0.50)	-0.03 (0.04)
No of minors	1.69 (1.11)	-0.026 (0.09)	1.79 (1.12)	0.075 (0.12)	1.73 (1.12)	-0.097 (0.12)
Own motorcycle	0.42 (0.49)	0.008 (0.04)	0.28 (0.45)	-0.035 (0.05)	0.42 (0.49)	0.003 (0.05)
Own car/truck	0.16 (0.37)	0.017 (0.03)	0.04 (0.19)	0.01 (0.03)	0.13 (0.34)	0.028 (0.04)
Own oxen/buffalo	0.55 (0.50)	0.032 (0.05)	0.39 (0.49)	0.109* (0.06)	0.53 (0.50)	0.033 (0.06)
Own pig	0.56 (0.50)	0.028 (0.04)	0.43 (0.50)	0.117** (0.06)	0.55 (0.50)	0.029 (0.05)
Own ox or buffalo cart	0.31 (0.46)	0.02 (0.04)	0.19 (0.40)	0.058 (0.05)	0.29 (0.45)	0.009 (0.05)
Hard roof	0.49 (0.50)	0.064 (0.04)	0.32 (0.47)	0.047 (0.05)	0.48 (0.50)	0.102** (0.05)
Hard wall	0.54 (0.50)	0.032 (0.04)	0.38 (0.49)	0.045 (0.06)	0.55 (0.50)	0.018 (0.05)
Hard floor	0.85 (0.36)	0.039 (0.03)	0.79 (0.41)	0.037 (0.05)	0.84 (0.37)	0.068* (0.04)
Have automatic toilet	0.07 (0.25)	-0.02 (0.02)	0.02 (0.13)	-0.01 (0.01)	0.05 (0.22)	0.005 (0.02)
Have pit toilet	0.12 (0.32)	0.018 (0.03)	0.11 (0.32)	0.02 (0.03)	0.13 (0.34)	0.001 (0.04)
Electricity	0.25 (0.43)	0.011 (0.04)	0.16 (0.37)	-0.01 (0.04)	0.23 (0.42)	-0.002 (0.05)
Pipe water	0.06 (0.24)	-0.001 (0.02)	0.03 (0.17)	-0.012 (0.01)	0.06 (0.23)	-0.013 (0.02)
Poverty Index (o to 292)	210.16 (60.18)	-1.609 (5.43)	245.13 (32.73)	-2.924 (5.14)	218.2 (51.66)	-11.771 (8.76)
Test score (0 to 25)	17.47 (4.81)	0.534 (0.52)	17.74 (4.71)	0.888 (0.68)	19.77 (3.22)	0.028 (0.48)
Number of students	940	2448	431	883	474	940
Number of schools	101	204	67	119	67	118

Columns (1), (3), (5): means and standard deviation, control group. Columns (2), (4), (6): difference with treatment, estimated by regressing each variable against corresponding treatment variable. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 2. Attrition and replacement households: baseline characteristics and differences

Variable	Attrition					Replacement Households		
	All baseline sample, no attritors (1)	Attritors (2)	Difference (2)-(1) (3)	Dif in dif		Original List (6)	Replacement households (7)	Difference (7)-(6) (8)
				Poverty (4)	Merit (5)			
Girl	0.5 '(0.50)	0.42 '(0.49)	-0.079*** '(0.02)	-0.103 '(0.07)	0.036 '(0.07)	0.51 '(0.50)	0.5 '(0.50)	-0.006 '(0.03)
No of minors	1.67 '(1.08)	1.64 '(1.11)	-0.033 '(0.05)	-0.068 '(0.17)	-0.173 '(0.17)	1.7 '(1.09)	1.64 '(1.06)	-0.052 '(0.06)
Own motorcycle	0.44 '(0.50)	0.38 '(0.48)	-0.064*** '(0.02)	0.062 '(0.07)	0.063 '(0.07)	0.42 '(0.49)	0.45 '(0.50)	0.024 '(0.03)
Own car/truck	0.2 '(0.40)	0.15 '(0.35)	-0.052*** '(0.02)	-0.019 '(0.03)	0.003 '(0.06)	0.19 '(0.39)	0.16 '(0.37)	-0.026 '(0.02)
Own oxen/buffalo	0.54 '(0.50)	0.49 '(0.50)	-0.049* '(0.03)	-0.134 '(0.09)	-0.059 '(0.08)	0.57 '(0.50)	0.53 '(0.50)	-0.044 '(0.03)
Own pig	0.53 '(0.50)	0.51 '(0.50)	-0.014 '(0.03)	-0.137 '(0.10)	-0.086 '(0.08)	0.57 '(0.50)	0.54 '(0.50)	-0.022 '(0.03)
Own ox or buffalo cart	0.32 '(0.47)	0.26 '(0.44)	-0.054** '(0.02)	-0.019 '(0.09)	0.017 '(0.07)	0.32 '(0.47)	0.3 '(0.46)	-0.022 '(0.03)
Hard roof	0.56 '(0.50)	0.44 '(0.50)	-0.119*** '(0.02)	0.093 '(0.08)	-0.011 '(0.08)	0.53 '(0.50)	0.55 '(0.50)	0.019 '(0.03)
Hard wall	0.59 '(0.49)	0.51 '(0.50)	-0.077*** '(0.02)	-0.023 '(0.09)	0.115 '(0.07)	0.56 '(0.50)	0.59 '(0.49)	0.023 '(0.03)
Hard floor	0.87 '(0.33)	0.84 '(0.37)	-0.038** '(0.02)	-0.043 '(0.06)	-0.053 '(0.05)	0.87 '(0.33)	0.87 '(0.34)	-0.003 '(0.02)
Have automatic toilet	0.1 '(0.30)	0.06 '(0.23)	-0.047** '(0.02)	0.053 '(0.04)	0.03 '(0.06)	0.06 '(0.24)	0.08 '(0.26)	0.014 '(0.01)
Have pit toilet	0.13 '(0.34)	0.11 '(0.32)	-0.018 '(0.01)	0.071 '(0.05)	-0.021 '(0.05)	0.12 '(0.33)	0.11 '(0.31)	-0.012 '(0.02)
Electricity	0.31 '(0.46)	0.22 '(0.41)	-0.085*** '(0.02)	0.048 '(0.07)	0 '(0.08)	0.25 '(0.44)	0.28 '(0.45)	0.025 '(0.03)
Pipe water	0.08 '(0.27)	0.07 '(0.25)	-0.016 '(0.01)	0.021 '(0.03)	0.042 '(0.04)	0.06 '(0.23)	0.07 '(0.26)	0.018 '(0.02)
Poverty Index (o to 292)	206.33 '(58.75)	215.52 '(56.94)	9.186*** '(2.56)	-0.508 '(7.84)	-0.208 '(10.48)	209.66 '(58.60)	211.19 '(57.23)	1.526 '(3.06)
Test score (0 to 25)	17.67 '(4.72)	17.7 '(4.91)	0.034 '(0.25)	-0.108 '(0.89)	0.002 '(0.59)	18.05 '(4.60)	18.16 '(4.74)	0.11 '(0.27)
Number of students	11088	1034	12122	2338	2448	3191	414	3605

Columns (1), (2), (6) and (7): means and standard deviation of each variable at baseline. Columns (3), (8): difference between groups. Columns (4) and (5), differences in differences, estimated by regressing each variable against corresponding treatment variable. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3. Impact on Enrollment and Attendance

	Reach Grade Six		Highest Grade Completed		Num. of hours in school, last 7 days (conditional on enrollment)	
	(1)	(2)	(1)	(2)	(1)	(2)
A. Poverty-based Treatment						
Treatment	0.186*** (0.04)	0.170*** (0.04)	0.349*** (0.11)	0.332*** (0.11)	3.466* (1.80)	2.865 '(1.87)
Control Mean	0.613*** (0.03)		5.377*** (0.09)		8.829*** (1.15)	
Covariates	✗	✓	✗	✓	✗	✓
No. Obs	883	883	831	831	665	665
F()	18.154	6.435	9.334	2.271	3.691	1.246
R2 Adj	0.042	0.18	0.025	0.145	0.015	0.131
B. Merit-based Treatment						
Treatment	0.131*** (0.05)	0.120*** (0.04)	0.234** (0.11)	0.182* (0.10)	1.374 (2.01)	0.635 '(1.55)
Control Mean	0.635*** (0.03)		5.448*** (0.08)		9.270*** (1.14)	
Covariates	✗	✓	✗	✓	✗	✓
No. Obs	940	940	897	897	713	713
F()	7.753	4.872	4.572	1.759	0.465	2.026
R2 Adj	0.02	0.155	0.011	0.122	0.002	0.199

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 4. Impact on test scores

	Mathematics		Digitspan Test	
	(1)	(2)	(1)	(2)
A. Poverty-based Treatment				
Treatment	-0.035 (0.09)	-0.041 (0.08)	-0.019 (0.09)	-0.059 (0.07)
Control Mean	0.018 (0.06)		0.015 (0.06)	
Covariates	✗	✓	✗	✓
No. Obs	883	883	883	883
F()	0.148	3.297	0.052	2.727
R2 Adj	0	0.167	0	0.113
B. Merit-based Treatment				
Treatment	0.15 (0.11)	0.170* (0.09)	0.147* (0.08)	0.149** (0.08)
Control Mean	0.165** (0.07)		0.081 (0.05)	
Covariates	✗	✓	✗	✓
No. Obs	940	940	940	940
F()	2.014	3.213	3.324	2.186
R2 Adj	0.005	0.16	0.006	0.093

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 5: Heterogeneity impacts on test scores by baseline test score and poverty index

	Mathematics		Digitspan	
	Test scores		Test scores	
	Low Achiever	High Achiever	Low Achiever	High Achiever
A. Poverty-based Treatment				
Treatment	-0.019 (0.11)	-0.069 (0.10)	-0.077 '(0.12)	-0.077 '(0.09)
Constant	3.444 (3.32)	-6.027* (3.58)	0.503 '(3.74)	-4.731 '(5.12)
Covariates	✓	✓	✓	✓
No. Obs	332	551	332	551
F()	2.143	2.351	2.932	4.171
R2 Adj	0.209	0.178	0.191	0.144
	Poverty index		Poverty index	
	Below median (non-poor)	Above median (poor)	Below median (non-poor)	Above median (poor)
B. Merit-based Treatment				
Treatment	0.053 (0.11)	0.233** (0.12)	0.194* '(0.11)	0.142 '(0.10)
Constant	1.803 (2.20)	1.872 (5.84)	-0.029 '(2.49)	-3.397 '(4.24)
Covariates	✓	✓	✓	✓
No. Obs	427	513	427	513
F()	2.754	4.46	2.324	4.465
R2 Adj	0.203	0.212	0.148	0.149

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 6. Impact of program on schoolwork (outside of school) and expenditures on education

	Student and Household (hh) effort					
	Time doing homework, studying or private lessons,			Total yearly expenditure in education, hh (US \$)		
	(1)	(2)	(3)	(1)	(2)	(3)
A. Poverty-based Treatment						
Treatment	0.443	0.42	0.134	-3.377	-0.925	-5.381
	(0.40)	(0.38)	(0.60)	'(3.44)	'(3.05)	'(5.45)
Above the median			-0.151			-1.559
			(0.59)			'(5.25)
Treatment X Above median			0.461			7.247
			(0.65)			'(6.74)
Constant	3.403***			21.216***		
	(0.29)			'(2.79)		
Covariates	✕	✓	✓	✕	✓	✓
No. Obs	599	599	599	622	622	622
F()	1.207	7.282	6.613	0.964	2.114	1.963
R2 Adj	0.003	0.145	0.146	0.002	0.125	0.128
B. Merit-based Treatment						
Treatment	0.614	0.700**	0.841*	3.867	5.815*	4.401
	(0.39)	(0.33)	(0.51)	'(3.46)	'(3.22)	'(4.56)
Above the median			-0.074			4.402
			(0.48)			'(4.58)
Treatment X Above median			-0.262			2.467
			(0.59)			'(6.28)
Constant	3.357***			20.518***		
	(0.25)			'(2.06)		
Covariates	✕	✓	✓	✕	✓	✓
No. Obs	657	657	657	684	684	684
F()	2.47	3.401	3.278	1.247	1.605	1.515
R2 Adj	0.005	0.176	0.177	0.002	0.154	0.157

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 7. Impact on Enrollment and Attendance on Nonrecipient Peers

	Reach Grade Six		Highest Grade Completed		Num. of hours in school, last 7 days (conditional on enrollment)	
	(1)	(2)	(1)	(2)	(1)	(2)
A. Poverty-based Treatment						
Treatment	0.086** (0.04)	0.082** (0.04)	0.108 -0.12	0.058 (0.11)	0.915 (2.04)	0.529 (1.82)
Control Mean	0.615*** (0.03)		5.376*** (0.08)		9.252*** (1.16)	
Covariates	✖	✓	✖	✓	✖	✓
No. Obs	785	785	732	732	576	576
F()	4.258	7.528	0.87	1.603	0.202	2.166
R2 Adj	0.008	0.172	0.002	0.125	0.001	0.191
B. Merit-based Treatment						
Treatment	-0.002 (0.06)	-0.009 (0.05)	-0.026 (0.13)	-0.099 (0.12)	1.418 (1.96)	0.291 (1.67)
Control Mean	0.584*** (0.03)		5.271*** (0.09)		8.694*** (1.14)	
Covariates	✖	✓	✖	✓	✖	✓
No. Obs	678	678	633	633	486	486
F()	0.001	6.94	0.038	1.765	0.523	3.195
R2 Adj	0	0.183	0	0.118	0.003	0.21

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.

Table 8. Impact on test scores on Nontreated peers

	Mathematics		Digitspan Test	
	(1)	(2)	(1)	(2)
A. Poverty-based Treatment				
Treatment	-0.034 (0.10)	-0.105 (0.08)	0.028 (0.09)	-0.009 (0.08)
Control Mean	-0.02 (0.06)		-0.017 (0.06)	
Covariates	✕	✓	✕	✓
No. Obs	785	785	785	785
F()	0.112	2.119	0.099	2.233
R2 Adj	0	0.153	0	0.115
B. Merit-based Treatment				
Treatment	0.006 (0.08)	-0.061 (0.08)	-0.013 (0.09)	-0.048 (0.08)
Control Mean	-0.229*** (0.06)		-0.113 (0.07)	
Covariates	✕	✓	✕	✓
No. Obs	678	678	678	678
F()	0.005	3.808	0.02	2.691
R2 Adj	0	0.129	0	0.131

Regression coefficient of dependent variable against treatment indicator. Column (1), without controls; Column (2), controlling for Table 1 baseline characteristics and district fixed effects. Robust standard errors clustered at the school level are presented in parentheses. ***, **, * indicates significance at the 1%, 5% and 10% levels respectively.