Long-Term Impacts of Alternative Approaches to Increase Schooling: Experimental Evidence from a Scholarship Program in Cambodia*

Felipe Barrera-Osorio
Harvard University

Andreas de Barros
Harvard University

Deon Filmer
The World Bank

February 28, 2019

1

Abstract

This randomized trial investigates the long-term effects of a primary school scholarship program in rural Cambodia. We estimate impacts—nine years after program inception—on educational attainment, cognitive skills, socioemotional outcomes, labor market outcomes, and well-being. Our results point to systematic improvements in attainment, but no average impacts on long-term cognitive or socioemotional outcomes. A merit-based (as opposed to poverty-based) targeting strategy did, however, increase cognitive outcomes, especially for poorer students. We also report positive effects on cognition for males. We find no improvements for labor market outcomes, yet positive effects on well-being, driven by recipients of the merit-based scholarships.

2

How does additional schooling impact long-term life outcomes? According to the canonical human capital model, labor markets remunerate the skills acquired during the education process (Becker 2009). According to a signaling model (Arrow 1973; Spence 1973), education provides the market with a signal of individuals' higher abilities; as a result, the market pays for these skills. Both models predict positive effects from investment in education. At the same time, emerging research is showing that, in many settings, increased schooling has not meant increased learning, which potentially limits the market returns to education (Pritchett 2013; The World Bank 2017). However, there are few empirical studies in low- and middle-income countries that isolate causal impacts of schooling on skills accumulation in the long run (Bouguen et al. 2018; Molina-Millan et al. 2016). Our aim is to contribute to this nascent literature in developing countries.[1] We present the causal long-term impacts of a scholarship program on cognitive skills, socioemotional outcomes, labor market outcomes, and socio-economic status and well-being[2], in a group of, on average, 21-year-old individuals who were offered the scholarship nine years earlier, in Cambodia.

Our study setup is the following. In 2008, 209 schools in Cambodia were randomly allocated between two treatment arms (104 schools) and a control group (105 schools). In approximately half of the treatment schools (51 schools), students in grade four received a scholarship based on merit—high-performing students were selected using a baseline test of math and language skills—and fourth-graders in the remaining treatment schools received a scholarship based on poverty—students were selected using a poverty index, based on household and family socio-economic characteristics. Scholarships were given to recipients for three years (i.e., until the completion of primary school), conditional on continued school participation and basic performance standards. A first follow-up study, three years after the program's

---

[1]Evaluations of Conditional Cash Transfer (CCT) or scholarship programs launched during the 1990s and the early 2000s in low- and medium-income countries are now allowing the exploration of such long-term effects of early-life interventions. The effects of these programs in the short run have been studied extensively; for reviews, see Baird et al. (2014); García and Saavedra (2017); Snilstveit et al. (2015).

[2]From this point on, we will refer to both socio-economic status and self-reported well-being as "well-being" for the sake of brevity.

inception, showed two main effects: higher school progression for individuals receiving either type of scholarship (compared with non-recipients), and impacts on cognitive outcomes (as measured by a math test and a test of working memory) only for those receiving a merit-based scholarship (Barrera-Osorio and Filmer 2016). In this paper we report impacts on a large set of long-term outcomes from data collected in late-2016/early-2017—nine years after the beginning of the program—from a sub-sample of the original study participants.

Our study contributes to an emerging literature on the causal, long-term effects of conditional cash transfers (CCTs) and scholarships. While programs vary in design, and contexts matter, the results are quite consistent that CCTs and scholarships, in the long-run, have positive effects on school progression (that is, non-treated individuals do not catch up with beneficiaries) and, in general, these programs increase formal education. These long-term effects were present in CCT programs in Nicaragua (Barham et al. 2013), Ecuador (Araujo et al. 2016), Colombia (Barrera-Osorio et al. forthcoming) and Mexico (Parker and Vogl 2018), as well as in a scholarship program in Ghana (Duflo et al. 2017). The effects on labor outcomes tend to be positive, but with some variation depending on the context and recipient gender. Positive effects were found in Nicaragua (ibid.), Kenya (Ozier 2016), and Mexico (ibid.), with heterogeneous effects by gender in Ghana (ibid.), and no effects in Ecuador (ibid.). Three of these long-term studies report delays in fertility and marriage: Ozier (2016) and Brudevold-Newman (2016), both in Kenya, and Duflo et al. (2017) in Ghana. Finally, two studies present causal evidence of positive effects of a scholarship program in Kenya on female empowerment and attitudes (Friedman et al. 2011; Jakiela et al. 2015). To the best of our knowledge, no studies have evidence on socioemotional measures and well-being outcomes (beyond health indicators).[3]

Our paper also contributes to the literature that studies how to build socioemotional (or "non-cognitive") outcomes. While evidence from high-income countries is increasing rapidly, research in low- and middle-income countries remains scant.[4] In

---

[3]Another strand of literature aims to measure the long-term effects of early childhood development interventions (for example, Walker et al. (2007) and Gertler et al. (2013)) and of youth training (Acevedo et al. 2016).

[4]See Heckman and Kautz (2014) and West et al. (2016) for an overview, and Deming (2017) for

these countries, even basic measurement issues are under-researched, with Laajaj and Macours (2017) being among the first to systematically test the cross-context validity of different constructs aimed at capturing socioemotional skills such as self-esteem, tenacity, conscientiousness, locus-of-control, or adaptability. They find that several of these measures are very noisy and that the cross-country validity of instruments cannot be taken for granted.[5] Beyond these basics, there is little analysis of how socioemotional skills are built, what role schooling might have in that process, or how the formation of cognitive and socioemotional skills might interact. The few analyses that have tried to shed light on these relationships are mainly descriptive (Claro et al. 2016; Kyllonen and Bertling 2013).

We make three main specific contributions to these literatures. First, we present impacts of the scholarship program on educational attainment, cognition, and socioemotional outcomes. We aim to document whether the initial short-term impacts on school progression are sustained. Likewise, we investigate if there is fade out of the initial impacts on cognitive skills that were detected three years after the program started.[6] In addition, ours is one of very few studies of impacts of schooling on socioemotional outcomes in a low-income country. Second, we present long-term effects of a scholarship program on labor market outcomes and well-being. Third, our experiment is unique as it allows for the analysis of varying targeting strategies (i.e., merit- versus poverty-based targeting).

Using the school randomization, we show intention-to-treat (ITT) results on "intermediate outcomes"—formal education, cognitive and socioemotional outcomes—

---

analyses of how the importance of social skills has grown in the U.S. labor market between 1980 and 2012. In high-income countries (and the United States, in particular), several studies have analyzed the production of socio-emotional skills in schools: Blazar (2017); Blazar and Kraft (2017); Jackson et al. (2014); Kraft (2017); Santorella (2017).

[5]A parallel study (Danon et al. 2018) is analyzing the properties of the various instruments that we use to measure of socioemotional skills. In this paper, we have retained only those measures that meet a basic threshold of psychometric properties

[6]There are several papers in high-income countries suggesting initial fade out from educational interventions (Bailey et al. 2017; Protzko 2015), and long-term effects on outcomes like health and criminal behavior (Anderson et al. 2009; Carneiro and Ginja 2014; Chetty et al. 2011, 2014; Currie and Thomas 2000; Deming 2009; Dynarski et al. 2013; Frisvold and Lumeng 2011; Garces et al. 2002; Heckman et al. 2010; Ludwig and Miller 2007).

and "long-term outcomes"—labor and well-being outcomes. In order to address potential problems of multiple hypothesis testing, we present results for "family indices"—that is indices that standardize the individual measures in each set of outcomes and calculate a weighted average (following Anderson (2008), by calculating inverse covariance matrix-weighted averages).[7] We have five main findings.

First, the results show that, despite some catch-up by the control group between 2011 and 2016, scholarship recipients have on average 0.241 more years of schooling than non-recipients. In comparison to their control group peers, treated individuals also improved their primary school completion rate by 8.0 percentage points, and had a 6.8 percentage point higher participation rate in formal education. The magnitudes of these impacts are in line with those for other programs that have attempted to reduce direct costs (for example scholarships; see Kremer et al. (2009), and Duflo et al. (2017)) and indirect costs of education (for example conditional cash transfers; see Fiszbein and Schady (2009)). These effects are slightly higher for poverty-based scholarship recipients than for merit-based recipients (though we cannot rule out that they are equal). Interestingly, the amount of the scholarship was extremely low (US$20 per year), implying a large price elasticity for education.

Second, we find positive effects on cognitive skills for merit-based scholarship receipts. On average, these students score 0.131 standard deviations higher in the cognitive family index than the control group, while we fail to reject equality to zero for the poverty-bases students. The effect is especially large for poor students who were offered a merit-based scholarship; the point estimate for the family index is 0.233 standard deviations (significant at the five-percent level). This result suggests that a merit-based scholarship does not necessarily increase inequalities.

Third, we do not find any systematic impacts on indicators related to two sets of measures of socioemotional outcomes: emotional and behavioral difficulties (as measured by the Strengths and Difficulties Questionnaire, "SDQ") and the "Big 5" personality traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism).[8] Impacts on these outcomes are systematically close to zero, and are

---

[7]We also report results for the individual underlying indicators.

[8]We collected information on other socioemotional outcomes such as grit and growth-mindset.

not statistically significant.

Fourth, we find a small (and marginally significant) effect of the program on the probability of working (2.6 percentage points, from a control mean of 91.9 percent) and a negative (but not statistically significant) effect on yearly earnings. This latter finding is puzzling as we report no negative impacts on labor market participation, the age of labor market entry, or the cognitive demands of respondents' occupations. We also document a positive, albeit not statistically significant, impact on recipients' daily reservation wage. It is important to recognize that our research—just as any of the few other long-term studies—measures effects when individuals are in their early-twenties, and as such it is perhaps too early to detect labor effects. We therefore interpret the results on labor outcomes with caution, since more time might make differences become more apparent–or even reverse the direction of impacts.[9]

Fifth, we find positive impacts on various measures of well-being. We report positive point estimates for all six indicators, and statistically significant impacts on recipients' perceived socio-economic status, quality of health, and quality of life. Per the family index, we find that the scholarship program improved recipients' well-being by 0.159 standard deviations (significant at the one-percent level). Merit-based scholarship recipients drive these positive results.

In addition to these main findings, we explore heterogeneity according to the scholarship recipient's gender.[10] We recognize that our study is weakly powered to detect differential effects; however, we find suggestive evidence that female recipients acquire more schooling than male ones (the effect on the education family index for males is 0.15 standard deviations, while the effect for females is 0.215 standard deviations; both of them statistically significant different from zero–but not statistically significant from each other). In contrast, we find positive effects on cognitive outcomes only for males (with a point estimate of 0.18 standard deviations; significant

_____

However, the psychometric and statistical properties of these measures in our context were weak (see Danon et al. 2018).

[9]Moreover, measurement issues (such as noisy data, seasonality, and informality) may especially complicate the ability to detect effects on labor market outcomes (Bouguen et al. 2018).

[10]Effects of the program might differ by gender either because of differential complementarities between skills and the demands in the labor market (Pitt et al. 2012) or because of differential costs and benefits in the decision of investing in education (Becker et al. 2010).

at the 10-percent level), and a negative point estimate of -0.07 standard deviations for females (the difference in effects is statistically significant, at the five-percent level). Similarly, we find that the program's positive effects on well-being outcomes are driven by impacts for males—the effect on the family index is of 0.203 standard deviations for males, and of 0.116 standard deviations for females respectively (only the former is statistically significant).[11] Together, these findings may suggest that, in the Cambodian context, females' additional educational investments are neither "rewarded" with increases in cognitive skill, nor with higher returns in the labor market or with improved well-being.

All in all, our results point towards significant effects from the scholarship on schooling, but effects on cognitive outcomes only for a certain group—merit-based scholarship recipients and especially for poor individuals among them—and no systematic effects on socioemotional outcomes. In turn, we do not find clear effects on labor market outcomes, but significant, positive effects on long-term well-being outcomes. These positive effects are driven by merit-based students, potentially indicating that the label attached to the scholarship matters. Our findings also reveal that male recipients benefited more from the scholarship than female recipients. Together, our results present a complex picture which nevertheless suggests that demand-side interventions, such as scholarships, and their particular targeting approaches can have important long-term effects.

The remainder of the paper proceeds as follows. In Section 1 we describe in more detail the study setup and context, in Sections 2, 3, and 4 we describe our estimation strategy, the study sample, and data, in Section 5 we present the results, and in Section 6 we provide concluding comments.

---

[11]We find a statistically significant difference in the effects, favoring males, on respondents' subjective social status, their asset ownership, and their perceived quality of health (see Table OA11, in the Online Appendix).

1   Intervention and Experimental Design

In 2008, the Government of Cambodia began implementing a new pilot scholarship program for grade four students in 209 public schools.[12] The program's stated goal was to reduce student drop-out rates and increase primary school completion, though the government also implicitly sought to improve students' educational performance. At the time, the program's 209 schools represented all public schools in three of the country's 25 provinces[13] (Mondulkiri, Ratanakiri, and Preah Vihear); the three provinces had been selected for having the highest drop-out rates in the upper primary grades (grades four to six), according to Cambodia's Education Management Information System (EMIS).[14] The program was phased in as a pilot over two years, with a random set of 104 schools starting in 2008/09 and the remaining schools entering in the following year (random assignment was stratified by province).

The scholarship program targeted students entering grade four, using one of two selection approaches. In a randomly selected half of the scholarship schools (51 schools), students were selected based on their combined performance on a test of Khmer and mathematics. This "merit-based" eligibility was determined through a centrally-scored test; the maximum possible score was 25. In the remaining 53 schools, they were selected based on a "poverty-based" approach. A student's "poverty score" was determined based on their self-reported (but validated) household and socio-economic characteristics; the poverty index ranges from 0 (richest household) to 292 (poorest household).[15] Under both approaches, half of a given school's fourth-graders qualified (i.e., the top half of performers, or the poorest half

---

[12]The program targeted 210 schools initially; here and elsewhere, we refer to those 209 schools that taught at least one grade-four student, at the beginning of the program.

[13]Here, we count the capital as Cambodia's 25th "province". More precisely, Phnom Penh is a special administrative district whose administrative characteristics partly resemble those of provinces.

[14]To limit the program's geographic scope, in Ratanakiri, only five of seven districts were selected, choosing those districts with the highest dropout rate. In the remaining two provinces, all districts were selected.

[15]The aptitude test was based on the 2005/06 Grade Three National Learning Assessment. The poverty assessment asked respondents about household demographics and possession of a list of assets (as provided in Table 2). See Barrera-Osorio and Filmer (2016) for more details on the student assessment and the poverty score.

of students).[16] Crucially (for our study), students in all 209 schools completed both types of assessments, independent of their school's assignment status.

Scholarships were offered to beneficiaries for three years (i.e., through the end of primary school), conditional on their continued enrollment, passing grades, and regular attendance. These requirements were moderately enforced.[17] Scholarships were disbursed as a lump-sum payment of USD20 in the first year, and two payments of USD10 in each of the following two years. As reported by Barrera-Osorio and Filmer (2016), these amounts represent about 3.3 percent of the yearly per capita expenditure in the study sample. These transfers are small compared to similar programs in other countries (Fiszbein and Schady 2009); even relatively small impacts may therefore be cost-effective.

Our experimental design exploits the randomized roll-out of the program over its two phases. In 2008/09, during phase one, fourth-graders in schools that were selected to disburse the program starting in the second phase did not receive any scholarship and did not become eligible in the years thereafter.[18] Note that a subset of these fourth-grade students would have been eligible under one of the two targeting schemes (merit-based or poverty-based), had their school been selected. In expectation, these two sub-samples are equal to their respective eligible peers from phase-one schools (below, we present supportive evidence that the two groups of students are in fact balanced, across phase-one and phase-two schools). Thus, we can identify the causal intention-to-treat effect of the scholarship program, under either of the two targeting approaches. As phase-one schools were moreover randomly assigned to either the poverty-based or merit-based targeting scheme, we can also compare the scholarship's effect across the two targeting schemes.

---

[16]Median students also qualified for the scholarship. The number of scholarships was determined using the previous year's official enrolment numbers.

[17]If a student lost her scholarship, its amount could not be re-allocated within the same school and the same year. Instead, the amount would be used for the subsequent cohort of fourth-graders.

[18]Recall that the program required students to maintain passing grades. Thus, a phase-one fourth-grader who attended a control group school could not become eligible in phase two by repeating the grade.

2   Estimation Framework

We estimate a generic production function model:

$$Y_{t,i} = \beta_0 + \beta_1 T_{0,i} + \gamma_1 \overline{m}_{0,i} + \gamma_2 \overline{p}_{0,i} + B\boldsymbol{X}_{0,i} + \mu_{t,i} \tag{1}$$

where Y are outcomes such as educational attainment, cognitive skills, socioemotional skills, labor outcomes, or measures of well-being (which include socio-economic status, SES). $\overline{m}$ and $\overline{p}$ are indicators for whether a student would *not* have qualified for a scholarship under the poverty-based (p) and merit-based (m) targeting-scheme, respectively. The inclusion of these control variables allows us to interpret the coefficient on T as the intention-to-treat effect of offering a scholarship. Vector $\boldsymbol{X}_{0,i}$ includes a rich set of baseline characteristics at the student's school-, village-, and individual-level (the next section describes these measures in greater detail). All estimations include province-level fixed effects and allow for the clustering of standard errors at the assignment level (i.e., within schools; cf. Abadie et al. (2017)). Equation 1 estimates an intention-to-treat model, with β_1 capturing the effect of offering a scholarship on outcomes Y, independent of the scholarship's targeting scheme.

We further assess the differential impacts of a merit-based vs. a poverty-based targeting approach in two ways, as shown in Equation 2 and Equation 3.

$$Y_{t,i} = \beta_0 + \beta_1 T_{0,i} + \beta_2 P_{0,i} + \gamma_1 \overline{m}_{0,i} + \gamma_2 \overline{p}_{0,i} + B\boldsymbol{X}_{0,i} + \mu_{t,i} \tag{2}$$

$$Y_{t,i} = \beta_0 + \beta_1 T_{0,i} + \beta_2 P_{0,i} + \beta_3 P\overline{m}_{0,i} + \beta_4 M\overline{p}_{0,i} + \gamma_1 \overline{m}_{0,i} + \gamma_2 \overline{p}_{0,i} + B\boldsymbol{X}_{0,i} + \mu_{t,i} \tag{3}$$

In Equation 2, we investigate whether effects differ across the two targeting schemes, from a policy perspective. If a social planner seeks to maximize (the aggregate of) treatment effects, then results for Equation 2 will inform her discrete choice for one of the two targeting approaches. Equation 2 adds an indicator P, denoting whether the treatment occurred in a school with the poverty-based targeting scheme. β_1 thus captures the intention-to-treat effect of the scholarship on outcome Y in a merit school, whereas the sum of β_1 and β_2 captures the intention-to-treat effect in a poverty school. Finally, the effect size of (and level of statistical significance for) β_2

11

indicates whether one targeting strategy dominates the other, from this perspective of maximizing effect sizes.

In turn, in Equation 3, we investigate whether effects differ across the two targeting schemes, *for individuals with comparable baseline characteristics*. Specifically, we compare treatment effects for students who would have qualified under either targeting approach. We would attribute a difference in treatment effects to the presentation of scholarships, or their mere "labelling" as either poverty-based or merit-based. In order to isolate this effect, Equation 3 adds interaction term $P\overline{m}$, denoting students in a school with the poverty-based targeting scheme, who would *not* have qualified under the merit-based scheme. Equation 3 also adds interaction term $M\overline{p}$, denoting students in a school with the merit-based targeting scheme, who would *not* have qualified under the poverty-based scheme. β_1 thus captures the intention-to-treat effect of the scholarship on outcome Y for poor, high-scoring students in a merit school, whereas the sum of β_1 and β_2 captures the intention-to-treat effect for poor, high-scoring students in a poverty school. Finally, the effect size of (and level of statistical significance for) β_2 indicates whether one labelling strategy dominates the other, for individuals with comparable baseline characteristics.

3 Sample and Internal Validity

In 2016, we followed a random sub-sample of 1,958 eligible students. This sample includes 577 eligible, high-scoring students in schools with the merit-based targeting strategy, 518 eligible, high-poverty students in schools with the poverty-based targeting strategy, and 863 students from control schools who would have qualified under at least one of the two targeting schemes.[19]

---

[19]Our overall sampling frame consisted of 3,918 eligible fourth-grade students (in the program's 209 schools teaching grade four at baseline), who participated in the baseline assessment, in December 2008 and January 2009. Of those, 1,908 respondents were randomly selected for the first three-year follow-up survey, in 2011. For this first follow-up, an additional 463 eligible "replacement" students were randomly selected, in case students could not to be found. In the 2016 follow-up, we tracked all students who had participated in the 2011 study and a random subset of 85 eligible respondents who had not been surveyed thus far (35 attritors from 2011 and 50 newly selected participants, from 2008/2009).

Following up a sample after almost 10 years is bound to be difficult, and this proved to be the case in this study. Overall, we have an attrition rate of 27.9 percent with a significant difference of 7.8 percentage points between treatment and control groups (see Table 1). Column (1) of Table 1 reports the coefficients from bivariate OLS regressions of attrition on baseline characteristics (in a model that includes province fixed effects to account for the study's stratified randomization). Males and better-off individuals are more likely to attrit. While the overall rate is somewhat higher than in other field studies[20], and there do seem to be some systematic correlates of attrition, we rule out that these correlates are systematically different across treatment and control groups. Column (6) reports the difference-in-difference among attritors and non-attritors, across treatment and control groups (again computed by OLS regression with province fixed effects). Only two out of 16 indicators—indicators of whether the respondent's household has a hard roof and owns pigs—is significant in this difference-in-difference specification. Notably, both reflect asset ownership, yet the coefficient's sign is positive for one and negative for the other. This finding is something we may expect in the case of testing multiple hypotheses. An additional test of the coefficients being jointly equal to zero, using seemingly unrelated estimation (SUR), yields a Chi-square statistic (and corresponding p-value) that does not allow us to reject that the two sub-samples are balanced.[21]

[Insert Table 1 here.]

In addition, and despite the different attrition rates between treatment and control groups, the characteristics of the treatment and control populations are balanced at baseline. Table 2 reports means for the treatment and control groups, and then the difference across these groups in a set of observable characteristics at baseline. None of the household characteristics are statistically significantly different across

---

[20]See Ghanem et al. (2018), who show that the median attrition rate in published field studies is around 15 percent——although this includes studies with much shorter follow-up times.

[21]In additional analyses, we investigate the two different targeting strategies, by treating them as separate treatment groups (see Appendix Table A1). SUR suggests that the coefficients may not be jointly equal to zero, when using separate samples. Further below we detail our robustness checks, which address this issue.

groups (even at the 10-percent level of significance). The only statistically significant difference at baseline between these groups is the indicator of being female: 57.3 percent of students in the treatment groups were girls, whereas 51.7 percent of students in the control group were. Notably, the samples are balanced on the two variables that were used to select scholarship recipients: the poverty index and the baseline test score. To further corroborate this finding (presented in Table 2), we also provide evidence for balance across the two different targeting strategies, in Appendix Table A2.

[Insert Table 2 here.]

Based on the findings discussed above, we conclude that the data generally support the fact that our experimental design is valid. First, the correlates of attrition are not different across treatment groups. Nevertheless, to account for the fact that the overall attrition rate differs across treatment and control groups, we perform a number of robustness checks using inverse probability weighting. We present these results in Appendix Tables A3 and A4. Second, we find that the treatment and control groups are balanced on observables at baseline for our study sample. Finally, to investigate whether the results might be driven by the (small) baseline difference in the gender balance across groups, we control for a gender indicator in all regressions and also provide an analysis of heterogeneous effects, by gender (in all specifications, we also control for a rich set of baseline variables).

4 Data and Measurement

Our analysis combines data from five main sources. First, we collect outcome data through in-person interviews at the respondents' residence, using handheld tablets. Second, to construct a variable reflecting intention-to-treat, we use the official government declaration ("Prakas") of scholarship recipients. Third, we match each respondent to baseline data—application forms and baseline tests—as collected in December 2008 and January 2009. Fourth, we construct a vector of control variables through administrative data on baseline school characteristics, as provided by the

country's Educational Management Information System (EMIS).[22] Fifth, we take advantage of the fact that Cambodia's 2008 census was conducted just before the scholarship program started. Using geographic coordinates, we match each school to its closest village and include this village's demographic characteristics as additional controls.[23]

Data collection for the baseline and three-year follow-up occurred from December 2008 to January 2009, and from May to September 2011, respectively. Data collection for our latest round of follow-up took from December 2016 to May 2017. We guaranteed data-quality by following standard monitoring procedures, as described by Glennerster (2017).[24]

The following discusses our newly collected outcome measures in greater detail. As education outcomes, we measure educational attainment (highest grade completed), whether a student completed primary education, and whether the respondent received any formal education since the early three-year follow-up (the latter two are binary variables).

We also collected data on four measures of cognitive skills. First, we administered a computer-adaptive math-test, in which respondents answered ten questions from a larger pool of 23 items.[25] We used a three-parameter logistic (3PL) item response theory (IRT) model with a single guessing parameter (Birnbaum 1968; Samejima

---

[22]We include a binary indicator of whether a school had access to drinking water, a binary indicator of whether the school had a toilet facility, the number of primary school classrooms, the number of newly enrolled fourth-graders, the number of teaching staff, and the school's income.

[23]Village-level data as published by the Cambodian National Institute of Statistics at the Ministry of Planning (2010). We control for the share of villagers who are literate in Khmer, the share of villagers with no schooling, the percentage of villagers engaged in crop or animal farming, the village's population size, and a continuous measure of villagers' household assets.

[24]First, during the first week of field work, we conducted 30 percent of re-surveys ("back-checks", usually within three days) and then reduced this number, for an overall back-check rate of 15.7 percent. Second, we spot-checked approximately 20 percent of interviews, provided immediate feedback, and offered repeat-trainings to enumerators. These spot-checks were not only conducted by field supervisors but also through additional, independent field-monitoring. Third, we ran daily analytics on newly collected data to spot irregularities, and to identify training needs. Finally, we employed 15 percent of staff as dedicated quality-control officers, such that steps to improve data quality could be taken immediately, as part of the regular data flow.

[25]To our best knowledge, this assessment constitutes the first computer-adaptive ability test as conducted during a household survey, in a developing country.

1969) to analyze responses to math tests from an evaluation of a similar scholarship program in Cambodia that was targeted to secondary school students (Filmer and Schady 2008). Participants in this assessment had been tested in two rounds, with overlapping items, and we follow the common (Stocking and Lord 1983) methodology for IRT-based scale equating.[26] Our adaptive test begins with the item of median difficulty. As the test is administered and respondents answer correctly or incorrectly, our assessment picks the next item to be displayed based on maximum information, re-calculates a respondent's ability estimate using expected a posteriori, and continues thereafter until ten items are administered for each respondent (cf. Bock and Mislevy 1982; van der Linden and Pashley 2010). The second assessment is a test of shapes and puzzles loosely based on the Raven's Progressive Matrices. This test is a measure of fluid intelligence; respondents are asked to complete 15 sets of pattern recognition.

Our third measure is a "Digit Span" test, which asks respondents to repeat sequences of single-digit numbers, of increasing length. This test is a common measure of respondents' working memory (Hamoudi and Sheridan 2015). Sequences are presented in sets of two and begin with two integers (asking respondents to repeat 2-1 and 1-3). No additional sequences are asked if a respondent fails to repeat both prompts; the last set of longest sequence presents two strings of eight integers (asking respondents to repeat 6-9-1-7-3-2-5-8 and 3-1-7-9-5-4-8-2).

The fourth outcome is a vocabulary test based on picture recognition, similar to a Peabody Picture Vocabulary Test (PPVT). This test asks respondents to identify the picture corresponding to a word which the enumerator reads out loud. For each word the respondent is asked to select from a choice of four pictures. The test is structured such that items become increasingly difficult (examples of easy items include, "citrus," and "garment"; items of highest difficulty include "vitreous" and "lugubrious"). A maximum of 96 items is presented in sets of 12, and no additional item is displayed if a respondent fails to answer at least five items correctly in a given set. The final skill estimate for each of the math, pattern recognition, and vocabulary recognition tests are calculated with a two-parameter logistic (2PL) IRT model. The

---

[26]We removed one item with low discrimination.

Digit Span test score reflects the number of integer sequences a respondent repeated correctly. All four measures are standardized (mean zero and standard deviation of one).

We report on two sets of socioemotional outcomes: we screen for emotional and behavioral difficulties with the Strengths and Difficulties Questionnaire ("SDQ"), and measure the "Big 5" personality traits. The SDQ represents a common screening instrument; we use (the official Khmer translation of) its most frequently used version with 25 items on psychological attributes (Goodman 1997). Following its scoring guidelines and official recommendations (*ibid.*), we report on three sub-scales, separated into 'internalizing problems' (emotional and peer symptoms, 10 items), 'externalizing problems' (conduct and hyperactivity symptoms, 10 items), and a scale of prosocial behavior (5 items).

To capture respondents' personality traits, the Big Five scale measures five core dimensions of personality. The five broad personality traits measured are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Evidence of the Big Five as being relevant (and associated with life outcomes) has been growing, beginning with the research of Fiske (1949) and later expanded upon by other researchers including Norman (1967), Smith (1967), Goldberg (1981), and McCrae and Costa (1987). We use the short 15-item Big Five Inventory (BFI-S) (Lang et al. 2011), with three items per personality trait. As with the indicators of cognitive skill, all measures of socioemotional outcomes are standardized.[27]

We also collected information on six labor market outcomes. We ask whether a respondent is currently working (yes or no) and the age at which she or he first started working. We also capture whether the individual participated in any formal or informal training that lasted at least one week, since 2011 (yes or no). We moreover construct a binary indicator of whether a respondent's main work activity is cognitively demanding. We categorize an occupation as such if it requires at least

---

[27]For further discussion on these measures, and their psychometric and statistical properties, see Danon et al. (2018). In addition to the measures we report on here, we collected data on respondents' level of grit (Duckworth and Quinn 2009) and their growth mindset (Dweck 2000). We do not report on results for these measures because of their poor psychometric properties in our data. The inclusion of either of these measures does not substantively change our results.

occasional use of reading, writing, mathematics, or a computer (according to the respondent). The survey also asked for respondents' income; our analysis reports on (the inverse hyperbolic sine of) yearly earnings and (the inverse hyperbolic sine of) a respondent's daily reservation wage, i.e., the minimum wage or payment for which a respondent is willing to accept work (both are reported in US dollars).[28]

Our last set of outcomes includes six indicators of socio-economic status and well-being. We assess subjective social status using a "MacArthur community ladder".[29] Respondents were shown a picture of a ladder with ten rungs and were told that higher rungs correspond to higher socio-economic status. They were then asked to place themselves on this ladder in relation to everyone in their community. As a second measure of socio-economic status, we construct an index of respondents' household assets, asking whether they possess items from a list similar to the one presented in Table 2. To calculate an individual's latent SES score, we borrow from the psychometric literature and estimate a two-parameter logistic (2PL) IRT model, placing responses from 2009, 2011, and 2016 on the same scale.[30] We also asked respondents to rate their satisfaction with life at present, all things considered, on a scale from one ("completely dissatisfied") to ten ("completely satisfied") and to rate their quality of life and health, respectively, on a scale from one ("poor") to five ("excellent"). The sixth and last measure screens for (minor) mental health disorders, using the General Health Questionnaire ("GHQ"). We use the short form of the questionnaire (GHQ-12) with Likert scoring (Goldberg and Williams 2006; Quek et al. 2001). All six measures are standardized (mean zero and standard deviation of one).[31]

---

[28]During the survey, respondents were allowed to answer either in US dollars (a currency commonly used in Cambodia) or in Riel (the local currency). The survey also allowed respondents to provide their answers in terms of varying payment modalities (including in-kind payments, piece-wise pay, and different payment frequencies, for example).

[29]For a description and bibliography of papers that use MacArthur ladders, see the MacArthur Foundation's Network on SES and Health website: http://www.macses.ucsf.edu/Research/Psychosocial/subjective.php.

[30]Filmer and Scott (2012) show that such an IRT approach produces similar household rankings when compared to other aggregation methods.

[31]We standardize by focusing on the endline measures for control group students (who would have qualified for at least one of the two types of scholarships, had they been in a treatment school

Finally, for each set of educational outcomes, cognitive outcomes, socioemotional outcomes, labor and SES and subjective well-being, we also calculate an overall "family index," following Anderson (2008).[32] These indices have the benefit of reducing the number of statistical tests (and the temptation to selectively focus on positive results). In constructing the indices, we ensured that the qualitative "direction" of the construct was preserved—higher values point to more desirable outcomes. However, our index construction is atheoretical and may therefore group together measurements with different underlying constructs. We therefore present and discuss results from both individual measurements and the family indices.

5 Results

We present and discuss our results in the sequence of our hypothesized "pathway model." First, we focus on whether the provision of a scholarship induced higher school accumulation, which in turn may have engendered cognitive and socioemotional skills. We denote these three broad outcomes—education progression, cognitive skills, and socioemotional outcomes—as "intermediate outcomes". We report impacts on these outcomes of offering a scholarship in Tables 3 to 5. We next turn to the question of whether the different types of scholarships induced different-sized impacts on their target populations. We call these "heterogeneous" effects. Because those target populations differ, we then turn to the question of whether the different types of scholarships had different impacts on the set of students who would have qualified for either type (but only received one type, depending on the school they were in). We call this effect the "labelling" effect (results for both heterogeneous and labelling effects are reported in Table 6). After these intermediate outcomes, we present the impacts of the scholarships on long-term outcomes: indicators of labor market outcomes (Table 7) and of the socioeconomic status and well-being of individuals (Table 8). We then discuss heterogeneous and labelling effects for these

---

instead).

[32]We also considered using an alternative index instead, following Kling et al. (2007). The alternative approach does not lead to qualitatively different conclusions.

outcomes (Table 9). Finally, we investigate whether treatment effects differ by gender (Table 10). All the estimates are intention-to-treat effects.

5.1 Intermediate Effects: Education, Cognitive and Socioemotional Outcomes

Table 3 presents impacts of the offer of a scholarship on the set of education outcomes, and on the family index for these outcomes. In these results we do not distinguish impacts by type of scholarship (as per Equation 1). Across all indicators, the scholarships had a consistently statistically positive impact on the acquisition of formal education. The treatment caused (1) an increase of highest grade completed of 0.241 grades compared to a control mean of 5.478 grades (a 4.4-percent increase); (2) an increase in the probability of completing primary school by 8.0 percentage points, compared to a control mean of 57.3 percent (a 14-percent increase); and (3) an increase in the probability of receiving any formal education during 2011-2017 of 6.8 percentage points, from a base of 75.4 percent (an increase of over nine percent).[33] Column (4) of Table 3 reports the program's impact on the family index of education outcomes: the increase is of 0.189 standard deviations.

[Insert Table 3 here.]

We now turn to the question of whether we detect any impacts on cognitive outcomes. Table 4 reports impacts on the computer-adaptive math test, the progressive matrices assessment, the Forward Digit Span test, and the picture recognition vocabulary test. None of the point estimates is statistically significantly different from

---

[33]An increase in 0.24 years of schooling is in the range of increases found as a consequence of much larger transfers and expenditures. For example, Behrman et al. (2013) report on a number of evaluations of the demand-side incentive Progresa program in Mexico which was associated with 0.5 additional grades after six years of the program for children who were aged six to eight (pre-program) in rural areas, and of 0.1 to 0.12 additional grades in urban areas for youth aged six to 20 (op. cit. Table 5.1). Progresa transferred to the typical grades 4, 5 and 6 student an average of (approximately) US$15 *per month*, in contrast to the Cambodia transfer of US$20 *per year*. Jackson et al. (2016) show that increasing per-student expenditure by ten percent in every year for all twelve years of public schooling in the US would causally lead to 0.31 more years of schooling attained.

zero. The estimate for the family index (reported in Column (5) of Table 4) is positive–reflecting the four positive point estimates of the individual indicators–but small (0.05 standard deviations) and not statistically significantly different from zero. From this we conclude that, overall, the offer of (any) scholarship (and the consequential increase in schooling) had no impacts on cognitive skills development that are detectable in the long-term.

[Insert Table 4 here.]

Table 5 reports the impact of scholarships on the various measures of socioemotional outcomes.[34] First, we present results for the Strengths and Difficulties Questionnaire (SDQ)—separating its three factors: Prosocial, internalizing, and externalizing behaviors. Second, we report results for the Big 5—separating its five factors: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The only statistically significant impact of the program is on the measure of neuroticism: the program led to a decrease in neuroticism by 0.110 standard deviations, significant at the ten-percent level. Given the lack of effects on the individual measures, it is perhaps unsurprising that we find no large or statistically significant impact on the socioemotional family index. The effect size is a small positive impact of 0.063 standard deviations but this estimate is not statistically significant (recall that for this index, all outcomes have been re-scaled such that higher values reflect more desirable outcomes). We therefore conclude that, overall, the offer of a scholarship had no detectable long-term impacts on socioemotional outcomes.

[Insert Table 5 here.]

In sum, these aggregate results point to systematic effects on school progression and acquisition of formal education, but no detectable impacts on long-term cognitive or socioemotional outcomes. Yet, these results could potentially mask heterogeneous impacts by type of scholarship, as was documented in the short-term follow-up analysis. The first three columns of Table 6 report the impacts of the scholarship offer

---

[34]We describe our theoretical model of the relationship between cognitive and socioemotional outcomes in Appendix B.

on the respective target population, by type of scholarship. The outcome variables are the family indices for education, cognition, and socioemotional outcomes.[35] The dummy variable structure is such that the coefficient of the overall treatment indicator ("Treatment (any)") captures the ITT effect of the merit-based scholarships on students who met the eligibility criterion for those scholarships; the coefficient on the treatment for poverty schools ("Treatment (poverty)") is the *additional* magnitude of the impact for the poverty-based scholarships on those who met the eligibility criterion for those scholarships (see Equation 2). Note that the statistical significance of this estimate is a test for the difference between impacts of merit versus poverty scholarships on each of their intended target populations. The sum of these two coefficients is the overall ITT impact of poverty-based scholarships on those who met the eligibility criterion for those scholarships (reported in the bottom panel of the table as "Treatment in Poverty Schools").

[Insert Table 6 here.]

The impacts on the education family index are positive for both merit- and poverty-based scholarship recipients. Students who were offered a merit-based scholarship scored 0.139 standard deviations higher on the family index, as compared to counterfactual students (Column 1). This impact is 0.106 standard deviation points higher for recipients of poverty-based scholarships, yielding a total impact of 0.245 standard deviations (while both of these point estimates are statistically significant, the difference between them is not, as indicated by the non-significance of the effect from "Treatment (poverty school)").[36]

The impacts on the cognition family index (Column 2) are statistically significant for the merit-based students (a 0.131 standard deviation estimate, significant

---

[35]Our interpretation of Tables 3 to 5 is that—per their rationale—each family index is indeed a good way to represent the group of impacts in each domain. We therefore report the effects using these indices; the results for the individual outcomes that make up the family indices (presented in the paper's Online Appendix), are consistent with those in Table 6.

[36]We speculate that differential impacts by type of scholarship on education acquisition could occur since poverty-based scholarship recipients have lower baseline education levels, and that it is easier to increase these from a lower level.

at the ten-percent level), and negative for the poverty-based treatment (-0.039, non-statitically significant). We can rule out that these effects are equal to each other. The results thus indicate that both types of scholarships likely increased the acquisition of schooling, even in the long-term, but only the merit-based induced effects on cognitive outcomes. The short-term differentials across scholarship types identified in the earlier assessment of this program (that is, positive effects on cognitive outcomes for recipients of merit-based scholarships) by Barrera-Osorio and Filmer (2016) were sustained over the long-term. In contrast, we don't see a clear pattern of heterogenity by type of scholarship on the socioemotional outcomes (Column 3).

These point estimates could potentially mask a different phenomenon: The target populations for the two scholarships types are different. As mentioned above, the poverty-based scholarships had slightly larger point estimates for education than merit-based ones, but this could be because the recipients were somewhat poorer and had lower baseline education indicators.[37] We address this potential effect by isolating the estimated effects to a subset of recipients who would have received a scholarship under either scheme (i.e., they are relatively high performing students from poor families at baseline), but depending on which school they were attending they would have (randomly) received a scholarship of one type or another. We call these effects "labelling" effects.[38] Columns (4) to (6) of Table 6 reports the results of this model. Here, the structure (per Equation 3) is such that the coefficient of the overall treatment indicator ("Treatment (any)") is the impact of being offered a merit-based scholarship for this sample of high-performing, high-poverty students. The additional impact of being offered a poverty-based scholarship for students with the same profile is captured by the coefficient on the dummy variable for schools in which poverty-based scholarships were distributed ("Treatment (poverty)"). The

---

[37]Note that this is not a problem with the experimental design: each of these impacts is derived from comparing a treatment group to a valid counterfactual group—both the treatment and counterfactual groups are different for the estimation of poverty-based and merit-based impacts.

[38]As discussed in Barrera-Osorio and Filmer (2016), differences according to this "labelling" could be driven by a number of factors, for example motivation associated with being called a "merit recipient," or discouragement associated with being called "poor," or being treated differently by teachers according to these "labels" or by a differential household educational investments across scholarship recipients.

statistical significance of this coefficient is a test of the difference in impacts for this specific subset of students. The sum of these two coefficients is the estimate of the impact of poverty-based scholarships, and is reported in the bottom panel of the table ("Effect in Poverty Schools, if high scores").[39]

The impact of being offered a merit-based scholarship on the education family index for high-scoring high-poverty students is 0.123 standard deviations. The impact of being offered a poverty-based scholarship for students with the same profile is 0.190 standard deviations. Of these two, only the latter is statistically significant. In addition, the test for these two effects being statistically significant cannot reject that they are the same. We attribute some of this pattern to the low power we might have—indeed, the additional effect of the poverty-based scholarship is not statistically significantly different from the scholarship-based one, even though the total effects are quite different. In sum, therefore we take these effects as indicating an overall gains in education, irrespective of the "label" attached to the scholarship.

In contrast, the coefficient for the family index for cognition outcomes is more consistently suggestive of labelling effects. The ITT effect of the offer of a merit-based scholarship on high-scoring, high-poverty students is 0.233 standard deviations (statistically significant, at the five-percent level).[40] The point estimate for a similar student—high-scoring, high-poverty—being offered a poverty-based scholarship is 0.036, which is not statistically significantly different from zero. We can reject that the point estimate is the same as the effect of merit-based scholarships. Labelling a poor student as "high-achiever" seems to have a positive impact on cognitive outcomes when combined with a scholarship, while the effect of labelling a high-achieving student as "poor" does not.

Finally, none of the results for socioemotional outcomes suggest much in the way of impacts. All of the effect sizes are small, and none are significantly different from

---

[39]These coefficients represent these effects because the model includes two additional dummy variables that capture the impacts of merit-based scholarships on non-poor recipients, and of poverty-based scholarships on low-scoring recipients, which therefore net out the impacts on those sub-populations (see Equation 3).

[40]We note that the impact of these scholarships for high-scoring low-poverty students is close to zero (and this difference is statistically significant).

each other. The overall lack of impacts discussed above do not seem to be masking any underlying heterogeneity.

5.2 Long-Term Effects: Labor and Well-Being Outcomes

The long-term effects of this program suggest two main findings so far: first, the program induced treated students to increase educational attainment (regardless of the type of scholarship); second, the program caused effects on cognitive outcomes only for merit-based scholarships receipts, especially for high-poverty, high-scoring merit-based students; third, the program did not seem to induce effects on socioemotional outcomes. The effects on education progression and cognitive outcomes are thus consistent with those identified in the earlier short-run follow-up analysis.

We now turn to our investigation of whether the program had impacts in other dimensions, namely labor market outcomes and measures of socioeconomic status and well-being. Table 7 reports the impacts on six labor outcomes—whether a respondent is currently working; the age the respondent first started working; whether the respondent received any work-related training (since 2011); a measure of the cognitive demands of work; the respondent's average yearly earnings; and the respondent's self-reported daily reservation wage.

The only significant effect of (any) scholarship on labor outcomes is on the probability of currently working, with a point estimate of 2.6 percentage points (in comparison to a control mean of 91.9 percent). All the other point estimates are close to zero, with the exception of a negative effects on yearly earnings, of -0.216 (inverse hyperbolic sine, USD).[41] As before, lack of statistical power may be affecting our ability to statistically detect impacts—an issue that is especially relevant for labor outcomes due to statistical noise, seasonal variability, and informality in the labor market (Bouguen et al. 2018). Overall, the coefficient of impact on the labor family index (Column 7) is small (0.025 standard deviations), and not statistically significant. Overall, we don't see a clear pattern of impacts on these outcomes.

---

[41]This results in predicted yearly earnings of USD965 for the control group, and USD708 in the treatment group.

[Insert Table 7 here.]

The impacts on well-being indicators are, however, statistically significant, including on the overall family index for this set of indicators (Table 8). Impacts on respondents' subjective social status and the quality of health and life in particular, are positive and statistically significant (effect sizes are, respectively, 0.176, 0.130 and 0.090 standard deviations). The point estimates of all the other measures point to "improvements"—point estimates for respondents' socioeconomic status index (based on household assets), reported life satisfaction, and quality of life are all positive; the coefficient for the index of health problems is negative. Together, the impact on the family index yields an effect size of 0.159 standard deviations, which is statistically significant (at the one-percent level). This suggests that while it is hard to identify the channels through which improvements in long-term well-being outcomes have operated, there nevertheless seem to be substantive impacts.

[Insert Table 8 here.]

Columns (1) and (2) of Table 9 report "heterogeneous" impacts on the family indices of labor outcomes and socioeconomic status/well-being by type of scholarship. These results suggest no difference in impacts by type of scholarship on the labor outcomes. In contrast, we detect impacts of merit-based scholarships on the socioeconomic status/well-being, but not for poverty-based scholarships: the effect size for the former is 0.232 standard deviations (significant at the one-percent level), whereas the effect size for poverty-based scholarships is 0.08 standard deviations (and not statistically significant). We can statistically reject that these effects are equal (the point estimate on the variable "Treatment (poverty school)" is -0.152, and statistically significant).

Columns (3) and (4) of Table 9 investigate whether part of these differential effects might come about because of labelling effects. Statistical power can play a role in the noise of these estimates since the cell sizes are becoming quite small (and the outcome measures are noisy).[42] The effect of the merit-based scholarships

---

[42]Here, we use the term "cell" to refer to either one of the four combinations of students' poverty- and merit-status.

on high-scoring high-poverty recipients on the labor family index (Column (3)) is 0.126 (not statistically significant), while the effect of poverty-based scholarships on similar recipients is close to zero (0.018, non-significant as well). The pattern of impacts for socioeconomic status / well-being on this sub-sample (poor and high achievers) are 0.190 for merit-based scholarships (statistically significant at the five-percent level) and close to zero for poverty-based ones. Still, we cannot rule out equality of coefficients.

In sum, we detect positive impacts on long-term measures of well-being, and these impacts appear to be driven by students receiving merit-based scholarships. The results on the labelling of the scholarships are consistent with impacts of merit scholarships and no impacts of poverty scholarships on poor and high-achieving students.

[Insert Table 9 here.]

5.3   Differential Effects by Gender

We last turn to another potential source of heterogeneity in program impacts, namely the gender of the scholarship recipient. There are a number of reasons why one might expect impacts to differ. For example, differences in early-child investments might mean that they have different skills at the time of the program, which would lead to differential impacts within education outcomes, or there may be different complementarities between skills and the demands in the labor market (Pitt et al. 2012), which could lead to different investments and outcomes. More generally, there may be gender-differentiated costs and benefits in the decision of investing in education (Becker et al. 2010).[43]

Table 10 reports impacts differentiated by gender. We estimate each model of the outcome of interest (five family indices) as a function of a dummy variable for having been offered a scholarship ("Treatment"), an indicator for being female ("Female"), and the interaction between these two. The structure is such that the coefficient of the

---

[43]For these analyses, it is important to keep in mind the slight imbalance across treatment and control groups in the gender ratio (see Table 2).

overall treatment variable is the effect on males, and the coefficient on the interaction term is the additional impact on females. The sum of these two coefficients is the overall impact on females, and is reported in the bottom part of the table ("Effect for females").

We find positive impacts on the education progression of females, with an effect size of 0.215 standard deviations (statistically significant). The point estimate for males is 0.152 (statistically significant as well), but we cannot rule out that the two estimates are equal. In contrast, we find differential impacts on cognition. The point estimate for males is 0.180, while the coefficient of the female-treatment interaction is -0.250 standard deviations lower (significant at the five-percent level). The total effect for females is -0.07 standard deviations (indistinguishable from zero). In addition, the impact on the family index for socioeconomic status/well-being is positive for males, with an effect size of 0.203 standard deviations, and the effect for females is 0.116 (non-statistically significant). We cannot statistically rule out that these are equal, but we nevertheless find suggestive evidence that the positive impact for males dominates the effect for females.[44] The point estimates for the other outcomes (family indices for socioemotional and labor outcomes) for females are all close to zero and not statistically significantly different from zero. It is striking that all the point estimates for males ("Treatment") are positive, while the point estimates of the interaction female-treatment are negative (with the exception of the education family index). Together, these findings suggest that, in the Cambodian context, educational investments are "rewarded" less strongly for females than for males.

[Insert Table 10 here.]

6 Conclusions

This study has investigated the long-term impacts of a scholarship program that increased schooling, with a particular focus on links between schooling, the devel-

---

[44]As reported in Table OA11 (in the Online Appendix), we find a statistically significant difference in the effects on respondents' subjective social status, their asset ownership, and their perceived quality of health, favoring males.

opment of cognitive and socioemotional skills, and labor market and well-being outcomes later in life. To this end, we evaluated the effects of a primary school scholarship program in rural Cambodia, nine years after the program's inception, tracking study participants when they were, on average, 21 years old.

Our results point to systematic impacts on school progression and acquisition of formal education, but no average impacts on long-term cognitive or socioemotional outcomes. We do identify positive impacts on cognitive skills among merit-based scholarship receipting, especially among poor (and high-achieving) students. While we find no positive impacts on average on labor market outcomes, we find that various measures of well-being have improved among recipients of merit-based scholarships. There is suggestive evidence of systematic gender differences in these long-term impacts.

We note two factors that are important to keep in mind when interpreting these results. First, they reflect the effect of increasing schooling by only about four additional months. While these may be critical months, inasmuch the program induced individuals to finish primary education (the effect is large: the scholarship induced an 8 percentage point increase in primary completion), it is possible that some of the key impacts of schooling on socioemotional skills happen early on (when both the control and treatment groups were still in school) or later in adolescence (when, for this population, both groups might have left school). Second, our relatively limited sample size may have reduced the precision of some estimates.

Our study has several implications. Prior research has argued that more schooling does not necessarily imply more learning (The World Bank 2017); in turn, our work highlights that more schooling—even if it enhances learning in the short run—may not lead to measurable improvements in socioemotional skills, and not necessarily to noticeable improvements in labor market outcomes among young adults. We note that it is possible that some of the positive effects might only manifest themselves later, as has been argued in the context of some long-run studies of early child development interventions (see discussion in Duncan and Magnuson (2013) for how this might explain findings from the United States). Indeed, our finding of positive impacts on long-term well-being suggests that there may be such unmeasured chan-

nels at work here. Importantly, our research also shows that targeting matters for cognitive outcomes. Nevertheless, additional research is needed in at least two main areas to better understand this puzzle.

First, our findings are consistent with research by Jackson (2018), suggesting that the school-based production of cognitive skills may not *necessarily* go hand-in-hand with improvements in socioemotional outcomes. However, research on how to purposefully foster socioemotional skill in school settings is only in its infancy, especially in low- and middle-income countries (see West et al. (2016) for an example from the United States). Moreover, much more research is needed to develop and validate robust measures of socioemotional skills that are appropriate for different cultural contexts (see Laajaj and Macours (2017) for an analysis and discussion of this challenge).

Second, our analysis of heterogeneous effects provides suggestive evidence that labor market effects may be concentrated among poorer beneficiaries, and among beneficiaries who are male. This result echoes the findings of Duflo et al. (2017), who find labor market effects for a subset of male students only. It will be important to understand how programs such as these can be designed in a way such that they also fully benefit female recipients.

Finally, we conclude with an implication for policy. The fact that it was poor recipients who received a merit-based scholarship that eventually had higher cognitive skills, and that it was merit-based scholarship recipients that ultimately reported better measures of well-being, suggests that there is not necessarily a trade-off between equity and efficiency when choosing how to target such a program. Of course this may be very context-specific (depending, for example, on a relatively low correlation between socioeconomic status and test scores at baseline), but it nevertheless suggests that there may be flexibility in how to design targeting schemes that reach both equity and efficiency objectives.

Tables

Table 1: Analysis of Attrition

	n (1)	OLS Coeff. (2)	Non-attritor T (3)	Attritor T (4)	Non-attritor C (5)	Attritor C (6)	Diff-in-Diffs Coeff. (7)
Female	1915	-0.099***	0.573	0.419	0.517	0.422	-0.053
		(0.022)	(0.495)	(0.494)	(0.500)	(0.495)	(0.055)
Number of minors	1879	0.007	1.74	1.672	1.713	1.849	-0.201
		(0.011)	(1.075)	(1.129)	(1.109)	(1.110)	(0.127)
Own motorcycle	1894	-0.027	0.328	0.29	0.376	0.363	-0.038
		(0.023)	(0.470)	(0.454)	(0.485)	(0.482)	(0.049)
Own car/truck	1877	-0.042	0.11	0.055	0.1	0.099	-0.043
		(0.041)	(0.313)	(0.229)	(0.300)	(0.299)	(0.033)
Own oxen/buffalo	1900	-0.077***	0.541	0.404	0.498	0.42	-0.035
		(0.027)	(0.499)	(0.492)	(0.500)	(0.494)	(0.059)
Own pig	1906	-0.048*	0.577	0.432	0.533	0.535	-0.132**
		(0.025)	(0.494)	(0.496)	(0.499)	(0.500)	(0.057)
Own ox or buffalo cart	1878	-0.083***	0.279	0.183	0.263	0.181	0.004
		(0.027)	(0.449)	(0.387)	(0.441)	(0.386)	(0.045)
Hard roof	1880	-0.087***	0.47	0.41	0.475	0.32	0.094*
		(0.023)	(0.499)	(0.493)	(0.500)	(0.467)	(0.056)
Hard wall	1908	-0.021	0.493	0.473	0.516	0.468	0.034
		(0.024)	(0.500)	(0.500)	(0.500)	(0.500)	(0.057)
Hard floor	1897	0.006	0.87	0.835	0.823	0.845	-0.048
		(0.034)	(0.336)	(0.372)	(0.382)	(0.363)	(0.041)
Have automatic toilet	1889	0.086	0.029	0.054	0.041	0.044	0.023
		(0.064)	(0.168)	(0.227)	(0.198)	(0.205)	(0.023)
Have pit toilet	1889	-0.01	0.124	0.151	0.117	0.091	0.048
		(0.039)	(0.329)	(0.358)	(0.322)	(0.288)	(0.040)
Electricity	1905	0.003	0.188	0.205	0.226	0.212	0.027
		(0.030)	(0.391)	(0.405)	(0.419)	(0.410)	(0.048)
Piped water	1901	0.022	0.029	0.046	0.042	0.033	0.025
		(0.068)	(0.167)	(0.209)	(0.201)	(0.179)	(0.023)
Poverty Index (0-292)	1958	0.000**	223.165	233.74	225.697	230.652	3.915
		(0.000)	(56.281)	(46.274)	(49.008)	(46.840)	(6.061)
Test score (0-25)	1958	-0.004	19.248	18.966	18.33	18.248	-0.299
		(0.003)	(3.929)	(4.300)	(4.348)	(4.432)	(0.536)
Joint significance: Ho: all coef. =0							
Chi-square							22.62
p-value							0.12
Attrition rate	1958			0.327		0.242	-0.078***
							(0.029)

Notes. All variables measured at baseline. Column (2) displays coefficients computed by bivariate OLS regressions of an attrition indicator on covariates, controlling for province fixed effects. Columns (3) to (6) display the means for the control group attritors, the treatment group attritors, the control group surveyed and the treatment group surveyed. Standard deviations in parentheses. Column (7) is the difference between the treatment group mean and the control group mean among attritors minus the difference between the treatment group mean and the control group mean among respondents. Differences in means are computed by OLS regression, controlling for province fixed effects. All standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. The Chi-square (and corresponding p-value below) is the result of a test testing for the individual coefficients being jointly equal to 0 using seemingly unrelated estimation.

Table 2: Balance at Baseline

	n (1)	All (2)	Treatment (3)	Control (4)	Difference (5)
Female	1378	0.55	0.573	0.517	0.049*
		(0.498)	(0.495)	(0.5)	(0.029)
Number of minors	1349	1.729	1.74	1.713	-0.028
		(1.089)	(1.075)	(1.109)	(0.101)
Own motorcycle	1362	0.348	0.328	0.376	-0.005
		(0.477)	(0.47)	(0.485)	(0.038)
Own car/truck	1351	0.106	0.11	0.1	0.025
		(0.308)	(0.313)	(0.3)	(0.025)
Own oxen/buffalo	1366	0.523	0.541	0.498	0.032
		(0.5)	(0.499)	(0.5)	(0.041)
Own pig	1372	0.559	0.577	0.533	0.038
		(0.497)	(0.494)	(0.499)	(0.041)
Own ox or buffalo cart	1351	0.272	0.279	0.263	0.007
		(0.445)	(0.449)	(0.441)	(0.036)
Hard roof	1349	0.472	0.47	0.475	0.021
		(0.499)	(0.499)	(0.5)	(0.041)
Hard wall	1372	0.502	0.493	0.516	0.004
		(0.5)	(0.5)	(0.5)	(0.044)
Hard floor	1366	0.851	0.87	0.823	0.035
		(0.357)	(0.336)	(0.382)	(0.033)
Have automatic toilet	1356	0.034	0.029	0.041	-0.006
		(0.181)	(0.168)	(0.198)	(0.015)
Have pit toilet	1356	0.121	0.124	0.117	0.016
		(0.326)	(0.329)	(0.322)	(0.03)
Electricity	1374	0.204	0.188	0.226	-0.012
		(0.403)	(0.391)	(0.419)	(0.035)
Piped water	1367	0.034	0.029	0.042	-0.005
		(0.182)	(0.167)	(0.201)	(0.015)
Poverty Index (0-292)	1411	224.208	223.165	225.697	-5.68
		(53.403)	(56.281)	(49.008)	(4.722)
Test score (0-25)	1411	18.87	19.248	18.33	0.383
		(4.13)	(3.929)	(4.348)	(0.46)
Joint significance: Ho: all coef. =0					
Chi-square					7.98
p-value					0.95

Notes. Minors refers to respondents age 14 and under; this may include the respondent. HH size refers to the number of people living in the respondent's household, including the respondent. Married is a dummy equal to 1 if the respondent is currently married and 0 if never married, divorced or separated. This variable is missing for minors. Currently working is a dummy equal to 1 if the respondent worked during the last week or has a job at the moment and 0 otherwise; respondents may work and also be a student. Column (1) presents the number of observations with endline information. Colums (2) to (4) display the means for all observations, the treatment group, and the control group, respectively. Standard deviations in parentheses. Column (5) is the difference between the treatment group mean and the control group mean. Differences in means are computed by OLS regression, controlling for province fixed effects. All standard errors in parentheses are clustered at the school level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The Chi-square (and corresponding p-value below) is the result of a test testing for the individual coefficients being jointly equal to 0, using seemingly unrelated estimation.

Table 3: Education Outcomes

	Highest grade completed	Completed primary	Received any formal education in 2011-2017	**Family index**
	(1)	(2)	(3)	**(4)**
Treatment	0.241**	0.080**	0.068**	**0.189***
	(0.114)	(0.032)	(0.028)	**(0.067)**
Observations	1,370	1,370	1,370	**1,370**
R-squared	0.110	0.105	0.087	**0.130**
F test	3.004	3.357	3.496	**4.058**
Covariates	Yes	Yes	Yes	**Yes**
Control mean	5.478	0.573	0.745	**0.039**

Notes. Estimated treatment effects. The dependent variable in column (1) is the highest grade the individual completed and is equal to -1 if the individual received no education, 0 if he only went to kindergarten and then ranges from 1 to 11 for Grade 1 to Grade 11. In column (2), the dependent variable is a dummy equal to 1 if the individual completed primary education. In column (3), the dependent variable is equal to 1 if the individual was enrolled in the formal education system during any of the years 2011 to 2016. In column (4), the family index is the inverse covariance matrix-weighted mean of the standardized dependent variables from the three previous columns following Anderson (2008). Treatment captures effects for students who received any treatment, under either scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: Cognitive Outcomes

	Math	Raven's	Forward Digit Span	Picture Recognition Vocabulary Test	**Family index**
	(1)	(2)	(3)	(4)	**(5)**
Treatment	0.062	0.091	-0.031	0.049	**0.050**
	(0.063)	(0.059)	(0.057)	(0.073)	**(0.065)**
Observations	1,370	1,370	1,369	1,370	**1,369**
R-squared	0.110	0.121	0.048	0.213	**0.150**
F test	6.558	8.035	2.320	9.674	**8.617**
Covariates	Yes	Yes	Yes	Yes	**Yes**
Control mean	0.023	0.007	0.010	0.031	**0.018**

Notes. Estimated treatment effects. The dependent variable in column (1) is the score on the mathematics computer adaptive test, computed using Item Response Theory (IRT) with a two parameter logistic (2PL) model, standardized. In column (2), the dependent variable is the score on the Raven's matrices test computed using IRT with a 2PL model, standardized. In column (3), the dependent variable is the standardized score on the digit span test using forward items only, standardized. In column (4), the dependent variable is the score on a Picture Recognition Vocabulary Test computed using IRT with a 2PL model, standardized. In column (5), the family index is the inverse covariance matrix-weighted mean of the standardized dependent variables from the four previous columns following Anderson (2008). Treatment captures effects for students who received any treatment, under either scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 5: Socioemotional Outcomes

	SDQ			Big 5					Family index
	Prosocial (1)	Internalizing (2)	Externalizing (3)	Openness (4)	Conscientiousness (5)	Extraversion (6)	Agreeableness (7)	Neuroticism (8)	(9)
Treatment	-0.001 (0.058)	-0.084 (0.059)	-0.007 (0.058)	0.030 (0.055)	-0.039 (0.056)	-0.035 (0.063)	-0.057 (0.058)	-0.110* (0.059)	0.063 (0.064)
Observations	1,367	1,368	1,368	1,368	1,369	1,368	1,368	1,370	1,360
R-squared	0.040	0.075	0.041	0.036	0.034	0.052	0.036	0.023	0.042
F-statistic	1.912	4.065	1.632	2.094	2.361	2.050	1.546	1.370	2.984
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	-0.014	-0.003	-0.004	0.009	0.001	0.016	0.000	0.022	-0.003

Notes. Estimated treatment effects. The dependent variable in column (1) is the score from 0 to 10 on the pro-social facet (the higher the score, the more pro-social) of the Strengths and Difficulties Questionnaire (SDQ), standardized. In column (2), the dependent variables is the score from 0 to 20 on the internalizing behavior facet (the higher the score, the more externalizing behavior problems) of the SDQ, standardized. In column (3), the dependent variables is the score from 0 to 20 on the externalizing behavior facet (the higher the score, the more externalizing behavior problems) of the SDQ, standardized. In columns (4) to (8), the dependent variables are the scores from 3 to 15 on the Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism facets of the Big Five scale, standardized. In column (9), the family index is the inverse covariance matrix-weighted mean of the standardized dependent variables from the eight first columns following Anderson (2008) (scores from columns (2), (3), and (8) have been flipped beforehand). Treatment captures effects for students who received any treatment, under either scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Heterogeneous and Labelling Effects on Schooling and Skills

	Heterogeneous Effects			Labelling Effects		
	Family index: Education	Family index: Cognition	Family index: Socioemotional	Family index: Education	Family index: Cognition	Family index: Socioemotional
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (any)	0.139*	0.131*	0.074	0.123	0.233**	0.126
	(0.077)	(0.078)	(0.078)	(0.105)	(0.104)	(0.113)
Treatment (poverty school)	0.106	-0.169*	-0.023	0.067	-0.197*	0.002
	(0.094)	(0.093)	(0.093)	(0.107)	(0.105)	(0.111)
Treatment (merit school and non-poor)				-0.008	-0.201	-0.067
				(0.126)	(0.141)	(0.151)
Treatment (poverty school and low scores)				0.162	-0.125	-0.195
				(0.110)	(0.141)	(0.158)
Observations	1,370	1,369	1,360	1,370	1,369	1,360
R-squared	0.132	0.153	0.042	0.133	0.155	0.043
Effect in Poverty Schools	0.245***	-0.039	0.051			
	(0.086)	(0.082)	(0.081)			
Effect in Poverty Schools, if high scores				0.190*	0.036	0.129
				(0.103)	(0.103)	(0.110)
F-statistic	4.283	8.402	2.920	4.338	8.128	2.920
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.039	0.018	-0.003	0.060	0.033	-0.082

Notes. Estimated treatment effects. The dependent variables in columns (1) to (3) (and columns (4) to (6), respectively) are the family indices from Tables 3, 4, and 5. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Labor Outcomes

	Currently working	Age started working	Any training since 2011	Cog. demands of main work (1/0)	Yearly earnings (inv. hyperbolic sine, USD)	Daily res. wage (inv. hyperbolic sine, USD)	Family index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.026*	0.172	-0.030	-0.014	-0.216	0.044	0.025
	(0.015)	(0.178)	(0.030)	(0.022)	(0.162)	(0.037)	(0.053)
Observations	1,300	1,327	1,370	1,299	1,334	1,353	1,241
R-squared	0.062	0.040	0.092	0.056		0.071	0.104
F-statistic	2.530	1.857	4.435	2.226	2.485	2.029	4.210
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.919	12.637	0.588	0.159	7.461	2.222	-0.001

Notes. Estimated treatment effects. The dependent variable in column (1) is a dummy equal to 1 if the individual is currently working, i.e. she worked for at least 1 hour during the last week or has a job at the moment but did not work during the last week. In column (2), the dependent variable is the age at which the individual started to work. In column (3), the dependent variable is a dummy equal to 1 if the individual participated in any formal or informal training that lasted at least one week, since 2011. In column (4), the dependent variable is a dummy equal to 1 if the main work activity demands cognitive ability (read, write, calculate, or use a computer) and 0 otherwise. In column (5), the dependent variable is the yearly earning expressed in US dollars and transformed using an inverse hyperbolic sine. In column (6), the dependent variable is the daily reservation wage in US dollars and transformed using an inverse hyperbolic sine. Treatment captures effects for students who received any treatment, under either scheme. In column (4), values for respondents who did not work have been imputed with 0, except if they were students. In columns (1) and (4), the sample is restricted to respondents who are not currently students. Column (2) includes everyone who ever worked. Column (4) includes only people who worked over the past 12 months. Columns (3) and (6) include the entire sample. In column (7), the family index is the inverse covariance matrix-weighted mean of the standardized dependent variables from the six first columns following Anderson (2008). Column (1), (2), (3), (4), (6), and (7) are estimated using OLS regression; column (5) is estimated using Tobit regression. All regressions control for province fixed effects, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Well-being Outcomes

	SES Ladder (village) (1)	SES Index (IRT) (2)	Life Satisfaction (3)	Quality of Health (4)	Quality of Life (5)	Health Issue Index (GHQ) (6)	Family index (7)
Treatment	0.176***	0.061	0.079	0.130**	0.090*	-0.046	0.159***
	(0.060)	(0.064)	(0.052)	(0.054)	(0.053)	(0.058)	(0.055)
Observations	1,370	1,370	1,369	1,370	1,370	1,355	1,354
R-squared	0.072	0.217	0.075	0.049	0.052	0.057	0.109
F-statistic	3.960	10.881	3.497	2.205	2.743	4.075	6.207
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	-0.014	0.002	-0.020	0.008	0.012	-0.011	-0.002

Notes. Estimated treatment effects. The dependent variable in column (1) is the score from 1 to 10 on an economic ladder as compared to people of the same age in the village, standardized. In column (2), the dependent variables is a socio-economic index constructed based on asset ownership computed using Item Response Theory with a two parameter logistic model, standardized. In column (3), the dependent variable is the score from 1 to 10 on a life satisfaction question, standardized. In column (4), the dependent variable is the score from 1 to 5 on a health quality question, standardized. In column (5), the dependent variable is the score from 1 to 5 on a life quality question, standardized. In column (6), the dependent variable is the standardized score on the General Health Questionnaire. In column (7), the family index is the inverse covariance matrix-weighted mean of the standardized dependent variables from all the previous columns following Anderson (2008). Treatment captures effects for students who received any treatment, under either scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Heterogeneous and Labelling Effects on Labor and Well-being Outcomes

	Heterogeneous Effects		Labelling Effects	
	Family index: Labor outcomes (1)	Family index: SES/Well-being (2)	Family index: Labor outcomes (3)	Family index: SES/Well-being (4)
Treatment (any)	0.031	0.232***	0.126	0.190**
	(0.062)	(0.070)	(0.090)	(0.090)
Treatment (poverty school)	-0.013	-0.152*	-0.080	-0.149
	(0.078)	(0.087)	(0.093)	(0.097)
Treatment (merit school and non-poor)			-0.221*	0.075
			(0.129)	(0.120)
Treatment (poverty school and low scores)			0.018	0.080
			(0.145)	(0.158)
Observations	1,241	1,354	1,241	1,354
R-squared	0.104	0.112	0.107	0.112
Effect in Poverty Schools	0.017	0.080		
	(0.069)	(0.071)		
Effect in Poverty Schools, if high scores			0.045	0.041
			(0.090)	(0.087)
F-statistic	4.106	6.104	4.080	5.949
Covariates	Yes	Yes	Yes	Yes
Control mean	-0.001	-0.002	-0.061	-0.088

Notes. Estimated treatment effects. The dependent variables in columns (1) and (2) (and columns (3) and (4), respectively) are the family indices from Tables 7 and 8. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

39

Table 10: Heterogeneous Treatment Effects by Gender

	Family index: Education (1)	Family index: Cognition (2)	Family index: Socioemotional (3)	Family index: Labor outcomes (4)	Family index: SES/Well-being (5)
Treatment	0.152**	0.180*	0.115	0.080	0.203***
	(0.075)	(0.093)	(0.084)	(0.075)	(0.072)
Female and treatment	0.063	-0.250**	-0.088	-0.123	-0.086
	(0.102)	(0.117)	(0.111)	(0.099)	(0.109)
Female	-0.093	-0.196*	-0.018	-0.258***	-0.151
	(0.102)	(0.114)	(0.102)	(0.089)	(0.099)
Observations	1,337	1,336	1,328	1,209	1,321
R-squared	0.132	0.154	0.046	0.106	0.109
Effect for females	0.215**	-0.070	0.027	-0.043	0.116
	(0.091)	(0.084)	(0.085)	(0.082)	(0.082)
F-statistic	4.302	9.145	3.246	4.147	5.971
Covariates	Yes	Yes	Yes	Yes	Yes
Control mean (females)	-0.015	-0.086	-0.041	-0.159	-0.140

Notes. Estimated treatment effects. The dependent variables in columns (1) to (5) are the family indices from Tables 3 to 8. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

40

References

Abadie, A., S. Athey, G. Imbens, and J. Wooldridge (2017, November). When Should You Adjust Standard Errors for Clustering? Technical Report w24003, National Bureau of Economic Research, Cambridge, MA.

Acevedo, P., G. Cruces, P. Gertler, and S. Martínez (2016, May). Soft Skills and Hard Skills in Youth Training Programs. Long Term Experimental Evidence from the Dominican Republic.

Anderson, K. H., J. E. Foster, and D. E. Frisvold (2009, February). Investing in Health: The Long-Term Impact of Head Start on Smoking. *Economic Inquiry 48*(3), 587–602.

Anderson, M. L. (2008, December). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Araujo, M. C., M. Bosch, and N. Schady (2016, September). Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap? Working Paper 22670, National Bureau of Economic Research.

Arrow, K. J. (1973, July). Higher Education as a Filter. *Journal of Public Economics 2*(3), 193–216.

Bailey, D., G. J. Duncan, C. L. Odgers, and W. Yu (2017, January). Persistence and Fadeout in the Impacts of Child and Adolescent Interventions. *Journal of Research on Educational Effectiveness 10*(1), 7–39.

Baird, S., F. H. Ferreira, B. Özler, and M. Woolcock (2014, January). Conditional, Unconditional and Everything in Between: A Systematic Review of the Effects of Cash Transfer Programmes on Schooling Outcomes. *Journal of Development Effectiveness 6*(1), 1–43.

Barham, T., K. Macours, and J. A. Maluccio (2013, May). Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *The American Economic Review 103*(3), 467–471.

Barrera-Osorio, F. and D. Filmer (2016, April). Incentivizing Schooling for Learning: Evidence on the Impact of Alternative Targeting Approaches. *Journal of Human Resources 51*(2), 461–499.

Barrera-Osorio, F., L. L. Linden, and J. E. Saavedra (forthcoming). Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia. *American Economic Journal: Applied Economics*.

Becker, G. S. (2009). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (3rd ed.). Chicago: University of Chicago Press.

Becker, G. S., W. H. J. Hubbard, and K. M. Murphy (2010, September). Explaining the Worldwide Boom in Higher Education of Women. *Journal of Human Capital 4*(3), 203–241.

Behrman, J. R., S. W. Parker, and P. E. Todd (2013). Incentives for Students and Parents. In P. Glewwe (Ed.), *Education Policy in Developing Countries*, pp. 137–192. Chicago: The University of Chicago Press.

Birnbaum, A. (1968). Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. *Statistical Theories of Mental Test Scores*.

Blazar, D. (2017, October). Validating Teacher Effects on Students' Attitudes and Behaviors: Evidence From Random Assignment of Teachers to Students. *Education Finance and Policy*, 1–52.

Blazar, D. and M. A. Kraft (2017, March). Teacher and Teaching Effects on Students' Attitudes and Behaviors. *Educational Evaluation and Policy Analysis 39*(1), 146–170.

Bock, R. D. and R. J. Mislevy (1982). Adaptive EAP Estimation of Ability in a Microcomputer Environment. *Applied psychological measurement 6*(4), 431–444.

Bouguen, A., Y. Huang, M. Kremer, and E. Miguel (2018, December). Using RCTs to Estimate Long-Run Impacts in Development Economics. Working Paper 25356, National Bureau of Economic Research, Cambridge, MA.

Brudevold-Newman, A. (2016, November). The Impacts of Free Secondary Education: Evidence from Kenya.

Carneiro, P. and R. Ginja (2014). Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy 6*(4), 135–173.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011, November). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics 126*(4), 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014, September). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review 104*(9), 2633–2679.

Claro, S., D. Paunesku, and C. S. Dweck (2016, August). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences 113*(31), 8664–8668.

Currie, J. and D. Thomas (2000). School Quality and the Longer-Term Effects of Head Start. *Journal of Human Resources 35*(4), 755–74.

Danon, A., A. de Barros, J. Das, and D. Filmer (2018, February). Measuring socioemotional skills in developing countries: Evidence from cambodia and pakistan.

Deming, D. (2009, June). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics 1*(3), 111–134.

Deming, D. (2017, November). The Growing Importance of Social Skills in the Labor Market. *The Quarterly Journal of Economics 132* (4), 1593–1640.

Duckworth, A. L. and P. D. Quinn (2009, February). Development and Validation of the Short Grit Scale (Grit–S). *Journal of Personality Assessment 91* (2), 166–174.

Duflo, E., P. Dupas, and M. Kremer (2017, April). The Impact of Free Secondary Education: Experimental Evidence from Ghana.

Duncan, G. J. and K. Magnuson (2013, February). Investing in Preschool Programs. *Journal of Economic Perspectives 27* (2), 109–132.

Dweck, C. S. (2000). *Self-theories: their role in motivation, personality, and development.* Essays in social psychology. Philadelphia, Pa.: Psychology Press. OCLC: 247496747.

Dynarski, S., J. Hyman, and D. W. Schanzenbach (2013, September). Experimental Evidence on the Effect of Childhood Investments on Postsecondary Attainment and Degree Completion. *Journal of Policy Analysis and Management 32* (4), 692–717.

Filmer, D. and N. Schady (2008, April). Getting Girls into School: Evidence from a Scholarship Program in Cambodia. *Economic Development and Cultural Change 56* (3), 581–617.

Filmer, D. and K. Scott (2012, February). Assessing Asset Indices. *Demography 49* (1), 359–392.

Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. *The Journal of Abnormal and Social Psychology 44* (3), 329–344.

Fiszbein, A. and N. R. Schady (2009, February). *Conditional Cash Transfers: Reducing Present and Future Poverty.* World Bank Publications. Google-Books-ID: aunlBU_2FsYC.

Friedman, W., M. Kremer, E. Miguel, and R. Thornton (2011, April). Education as Liberation? Technical Report w16939, National Bureau of Economic Research, Cambridge, MA.

Frisvold, D. E. and J. C. Lumeng (2011, March). Expanding Exposure Can Increasing the Daily Duration of Head Start Reduce Childhood Obesity? *Journal of Human Resources 46*(2), 373–402.

Garces, E., D. Thomas, and J. Currie (2002). Longer-Term Effects of Head Start. *The American Economic Review 92*(4), 999–1012.

García, S. and J. E. Saavedra (2017, October). Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis. *Review of Educational Research 87*(5), 921–965.

Gertler, P., J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S. Chang, and S. Grantham-McGregor (2013, June). Labor Market Returns to Early Childhood Stimulation: A 20-year Followup to an Experimental Intervention in Jamaica. Technical Report w19185, National Bureau of Economic Research, Cambridge, MA.

Ghanem, D., S. Hirshleifer, and K. Ortiz-Becerra (2018). Testing Attrition Bias in Field Experiments.

Glennerster, R. (2017). The Practicalities of Running Randomized Evaluations: Partnerships, Measurement, Ethics, and Transparency. In A. V. Banerjee and E. Duflo (Eds.), *Handbook of Economic Field Experiments*, Volume 1, pp. 175–243. Elsevier.

Goldberg, D. and P. Williams (2006). *A user's guide to the General Health Questionnaire*. GL assessment.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of personality and social psychology 2*(1), 141–165.

Goodman, R. (1997, July). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry 38*(5), 581–586.

Hamoudi, A. and M. Sheridan (2015, November). Unpacking the Black Box of Cognitive Ability. A novel tool for assessment in a population based survey.

Heckman, J. J. and T. Kautz (2014, January). Fostering and Measuring Skills: Interventions that Improve Character and Cognition. In J. J. Heckman, J. E. Humphries, and T. Kautz (Eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, pp. 341–430. Chicago: University of Chicago Press. Google-Books-ID: gJGPAgAAQBAJ.

Heckman, J. J., S. H. Moon, R. Pinto, P. Savelyev, and A. Yavitz (2010). A New Cost-Benefit and Rate of Return Analysis for the Perry Preschool Program: A Summary. In A. J. Reynolds, A. J. Rolnick, M. M. Englund, and J. A. Temple (Eds.), *Childhood Programs and Practices in the First Decade of Life*, pp. 366–380. Cambridge: Cambridge University Press.

Jackson, C. K. (2018, June). What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*.

Jackson, C. K., R. C. Johnson, and C. Persico (2016, February). The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. *The Quarterly Journal of Economics 131*(1), 157–218.

Jackson, C. K., J. E. Rockoff, and D. O. Staiger (2014). Teacher Effects and Teacher-Related Policies. *Annual Review of Economics 6*(1), 801–825.

Jakiela, P., E. Miguel, and V. L. te Velde (2015, September). You've earned it: estimating the impact of human capital on social preferences. *Experimental Economics 18*(3), 385–407.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007, January). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

46

Kraft, M. A. (2017). Teacher Effects on Complex Cognitive Skills and Social-Emotional Competencies. *Journal of Human Resources*, 0916–8265R3.

Kremer, M., E. Miguel, and R. Thornton (2009, July). Incentives to Learn. *The Review of Economics and Statistics 91*(3), 437–456.

Kyllonen, P. C. and J. P. Bertling (2013). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, and D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, pp. 277–285. London: Chapman & Hall.

Laajaj, R. and K. Macours (2017, March). Measuring skills in developing countries. Working Paper WPS8000, The World Bank, Washington, D.C.

Lang, F. R., D. John, O. Lüdtke, J. Schupp, and G. G. Wagner (2011, June). Short assessment of the Big Five: robust across survey methods except telephone interviewing. *Behavior Research Methods 43*(2), 548–567.

Ludwig, J. and D. L. Miller (2007, February). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *The Quarterly Journal of Economics 122*(1), 159–208.

McCrae, R. R. and P. T. Costa (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology 52*(1), 81–90.

Molina-Millan, T., T. Barham, K. Macours, J. A. Maluccio, and M. Stampini (2016, October). Long-term Impacts of Conditional Cash Transfers in Latin America: Review of the Evidence. Working Paper IDB-WP-732, Inter-American Development Bank, Washington, D.C.

National Institute of Statistics, Ministry of Planning (2010). 2008 Census CambodiaRedatam+SP.

Norman, W. T. (1967, April). 2800 personality trait descriptors - normative operating characteristics for a university population. Technical Report UM-00310-1-T, Michigan University, Ann Arbor.

Ozier, O. (2016, December). The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis. *Journal of Human Resources*.

Parker, S. W. and T. Vogl (2018, February). Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico. Working Paper 24303, National Bureau of Economic Research.

Pitt, M. M., M. R. Rosenzweig, and M. N. Hassan (2012, December). Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy. *American Economic Review 102*(7), 3531–3560.

Pritchett, L. (2013). *The Rebirth of Education: Schooling Ain't Learning*. Washington, D.C.: Brookings Institution Press for Center for Global Development.

Protzko, J. (2015, November). The Environment in Raising Early Intelligence: A Meta-Analysis of the Fadeout Effect. *Intelligence 53*, 202–210.

Quek, K. F., W. Y. Low, A. H. Razack, and C. S. Loh (2001, October). Reliability and validity of the General Health Questionnaire (GHQ-12) among urological patients: A Malaysian study. *Psychiatry and Clinical Neurosciences 55*(5), 509–513.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores. *Psychometrika 34*(4), Part 2.

Santorella, E. (2017, November). Multi-Dimensional Teacher Effects.

Smith, G. M. (1967, March). Personality correlates of cigarette smoking in students of college age. *Annals of the New York Academy of Sciences 142*(1), 308–321.

Snilstveit, B., J. Stevenson, D. Phillips, M. Vojtkova, E. Gallagher, T. Schmidt, H. Jobse, M. Geelen, M. G. Pastorello, and J. Eyers (2015). Interventions for improving learning outcomes and access to education in low-and middle-income

countries: a systematic review. Technical report, International Initiative for Impact Evaluation, London.

Spence, M. (1973, August). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374.

Stocking, M. L. and F. M. Lord (1983, April). Developing a Common Metric in Item Response Theory. *Applied Psychological Measurement 7*(2), 201–210.

The World Bank (2017). *Learning to Realize Education's Promise. World Development Report 2018*. Washington, D.C.: The World Bank. OCLC: 992735784.

van der Linden, W. J. and P. J. Pashley (2010). Item Selection and Ability Estimation in Adaptive Testing. In W. J. van der Linden and C. A. W. Glas (Eds.), *Elements of adaptive testing*, Statistics for social and behavioral sciences, pp. 3–30. New York: Springer. OCLC: ocn465370134.

Walker, S. P., S. M. Chang, C. A. Powell, E. Simonoff, and S. M. Grantham-McGregor (2007, November). Early Childhood Stunting Is Associated with Poor Psychological Functioning in Late Adolescence and Effects Are Reduced by Psychosocial Stimulation. *The Journal of Nutrition 137*(11), 2464–2469.

West, M. R., M. A. Kraft, A. S. Finn, R. E. Martin, A. L. Duckworth, C. F. O. Gabrieli, and J. D. E. Gabrieli (2016, March). Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling. *Educational Evaluation and Policy Analysis 38*(1), 148–170.

Appendix A   Additional tables

Table A1: Analysis of Attrition, by Targeting Approach

	Merit Scholarship					Poverty Scholarship				
	Attritor C	Attritor T	Non-attritor C	Non-attritor T	Diff-in-Diffs	Attritor C	Attritor T	Non-attritor C	Non-attritor T	Diff-in-Diffs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Female	0.452	0.391	0.487	0.491	-0.062	0.485	0.402	0.518	0.632	-.184**
	(.499)	(.49)	(.501)	(.501)	(.072)	(.501)	(.492)	(.5)	(.483)	(.082)
Number of minors	1.831	1.527	1.739	1.631	-0.184	1.899	1.83	1.81	1.864	-0.149
	(1.132)	(1.09)	(1.124)	(1.091)	(.165)	(1.111)	(1.154)	(1.117)	(1.038)	(.188)
Own motorcycle	0.373	0.355	0.44	0.424	-0.001	0.266	0.232	0.274	0.227	-0.014
	(.485)	(.481)	(.497)	(.495)	(.075)	(.443)	(.424)	(.447)	(.419)	(.063)
Own car/truck	0.114	0.082	0.117	0.174	-0.08	0.024	0.027	0.024	0.042	-0.008
	(.319)	(.275)	(.322)	(.38)	(.05)	(.152)	(.162)	(.153)	(.202)	(.026)
Own oxen/buffalo	0.458	0.427	0.543	0.587	-0.058	0.361	0.384	0.381	0.487	-0.048
	(.5)	(.497)	(.499)	(.493)	(.079)	(.482)	(.489)	(.486)	(.501)	(.074)
Own pig	0.554	0.427	0.545	0.619	-.198***	0.479	0.446	0.432	0.535	-0.11
	(.499)	(.497)	(.499)	(.486)	(.072)	(.501)	(.499)	(.496)	(.499)	(.078)
Own ox or buffalo cart	0.181	0.182	0.284	0.302	0	0.16	0.188	0.179	0.227	-0.003
	(.386)	(.387)	(.452)	(.46)	(.058)	(.367)	(.392)	(.384)	(.419)	(.063)
Hard roof	0.367	0.464	0.507	0.599	0.011	0.231	0.366	0.363	0.334	.159**
	(.484)	(.501)	(.501)	(.491)	(.078)	(.423)	(.484)	(.482)	(.472)	(.078)
Hard wall	0.518	0.573	0.575	0.581	0.052	0.402	0.357	0.393	0.391	-0.037
	(.501)	(.497)	(.495)	(.494)	(.071)	(.492)	(.481)	(.489)	(.489)	(.079)
Hard floor	0.831	0.827	0.845	0.933	-.087*	0.805	0.866	0.774	0.816	0.041
	(.376)	(.38)	(.363)	(.25)	(.05)	(.398)	(.342)	(.419)	(.388)	(.058)
Have automatic toilet	0.048	0.055	0.053	0.052	0.006	0.012	0.036	0.009	0	0.033
	(.215)	(.228)	(.224)	(.223)	(.035)	(.108)	(.186)	(.094)	(0)	(.024)
Have pit toilet	0.102	0.118	0.129	0.134	0.011	0.083	0.17	0.113	0.096	.097*
	(.304)	(.324)	(.336)	(.341)	(.056)	(.276)	(.377)	(.317)	(.295)	(.055)
Electricity	0.235	0.227	0.232	0.224	-0.002	0.166	0.179	0.161	0.142	0.02
	(.425)	(.421)	(.423)	(.417)	(.078)	(.373)	(.385)	(.368)	(.349)	(.051)
Piped water	0.03	0.045	0.041	0.047	0.009	0.006	0.027	0.018	0.011	0.026
	(.171)	(.209)	(.199)	(.211)	(.031)	(.077)	(.162)	(.133)	(.106)	(.018)
Poverty Index (0-292)	234.56	233.855	227.32	220.041	5.294	251.503	248.875	249.854	247.7	-1.834
	(39.843)	(39.426)	(43.884)	(45.032)	(6.966)	(24.17)	(26.574)	(27.738)	(27.71)	(4.551)
Test score (0-25)	19.855	19.791	19.821	19.93	-0.226	17.929	17.973	17.869	18.827	-1.095
	(3.281)	(3.454)	(3.15)	(2.851)	(.535)	(4.501)	(5.16)	(4.65)	(4.745)	(.817)
Observations	166	110	341	344	961	169	112	336	353	970
Attrition rate					0.287					0.29
Joint significance: Ho: all coef. =0										
Chi-square					26.09					24
p-value					0.05					0.09

Notes. All variables measured at baseline. Columns (1) to (4) (and (6) to (9), respectively) display the means for the control group attritors, the treatment group attritors, the control group surveyed and the treatment group surveyed. Standard deviations in parentheses. Column (5) and Column (10) provide the difference between the treatment group mean and the control group mean among attritors minus the difference between the treatment group mean and the control group mean among respondents. Differences in means are computed by OLS regression, controlling for province fixed effects. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. The Chi-square (and corresponding p-value below) is the result of a test testing for the individual coefficients being jointly equal to 0 using seemingly unrelated estimation.

Table A2: Balance at Baseline, by Targeting Approach

| | | Merit Scholarship | | | | | Poverty Scholarship | | | |
	n (1)	All (2)	Treatment (3)	Control (4)	Difference (5)	n (6)	All (7)	Treatment (8)	Control (9)	Difference (10)
Female	825	0.503 (0.5)	0.506 (0.501)	0.5 (0.501)	0.003 (0.035)	780	0.59 (0.492)	0.642 (0.48)	0.535 (0.499)	0.093*** (0.035)
Number of minors	804	1.68 (1.109)	1.623 (1.091)	1.738 (1.124)	-0.132 (0.123)	765	1.822 (1.084)	1.861 (1.046)	1.782 (1.122)	0.081 (0.119)
Own motorcycle	813	0.424 (0.495)	0.428 (0.495)	0.421 (0.494)	0.014 (0.054)	776	0.246 (0.431)	0.225 (0.418)	0.268 (0.443)	-0.024 (0.042)
Own car/truck	799	0.148 (0.355)	0.171 (0.377)	0.124 (0.33)	0.028 (0.035)	780	0.037 (0.189)	0.048 (0.215)	0.026 (0.159)	0.013 (0.022)
Own oxen/buffalo	817	0.558 (0.497)	0.581 (0.494)	0.535 (0.499)	0.014 (0.056)	776	0.441 (0.497)	0.5 (0.501)	0.38 (0.486)	0.065 (0.051)
Own pig	822	0.585 (0.493)	0.609 (0.489)	0.562 (0.497)	0.031 (0.049)	776	0.488 (0.5)	0.545 (0.499)	0.431 (0.496)	0.079 (0.055)
Own ox or buffalo cart	808	0.298 (0.458)	0.313 (0.464)	0.284 (0.451)	-0.001 (0.045)	767	0.211 (0.408)	0.244 (0.43)	0.177 (0.382)	0.032 (0.045)
Hard roof	806	0.548 (0.498)	0.583 (0.494)	0.514 (0.5)	0.062 (0.05)	763	0.359 (0.48)	0.354 (0.479)	0.365 (0.482)	-0.024 (0.052)
Hard wall	821	0.568 (0.496)	0.566 (0.496)	0.569 (0.496)	-0.011 (0.052)	776	0.403 (0.491)	0.416 (0.494)	0.39 (0.488)	0.011 (0.061)
Hard floor	811	0.882 (0.323)	0.921 (0.27)	0.842 (0.365)	0.063* (0.033)	774	0.801 (0.399)	0.818 (0.386)	0.783 (0.413)	0.002 (0.047)
Have automatic toilet	804	0.052 (0.223)	0.055 (0.229)	0.049 (0.217)	0.003 (0.023)	775	0.008 (0.088)	0.003 (0.05)	0.013 (0.114)	-0.01 (0.009)
Have pit toilet	804	0.132 (0.339)	0.131 (0.337)	0.133 (0.34)	0.005 (0.039)	775	0.112 (0.316)	0.116 (0.321)	0.108 (0.311)	0.018 (0.034)
Electricity	822	0.232 (0.423)	0.227 (0.419)	0.238 (0.426)	-0.008 (0.047)	777	0.16 (0.366)	0.148 (0.355)	0.172 (0.378)	-0.022 (0.04)
Piped water	818	0.045 (0.208)	0.042 (0.2)	0.049 (0.215)	-0.005 (0.021)	774	0.019 (0.138)	0.015 (0.123)	0.023 (0.152)	-0.01 (0.013)
Poverty Index (0-292)	855	211.519 (60.088)	204.589 (66.272)	218.63 (52.127)	-10.969 (8.114)	784	245.043 (31.622)	243.426 (32.419)	246.703 (30.736)	-0.5 (4.345)
Test score (0-25)	855	19.773 (3.068)	19.818 (2.867)	19.727 (3.263)	0.205 (0.472)	784	18.221 (4.735)	18.627 (4.754)	17.804 (4.685)	1.002 (0.667)

Joint significance: Ho: all coef. =0
Chi-square					14.04					15.25
p-value					0.6					0.51

Notes. Column (1) and Column (6) present the number of observations in the analysis sample (excluding observations with imputed baseline information). Colums (2) to (4) (and (7) to (9), respectively) display the means for the full sample, the treatment group, and the control group, respectively. Standard deviations in parentheses. Column (5) and Column (10) provide the difference between the treatment group mean and the control group mean. Differences in means are computed by OLS regression, controlling for province fixed effects. Standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. The Chi-square (and corresponding p-value below) is the result of a test testing for the individual coefficients being jointly equal to 0 using seemingly unrelated estimation.

52

Table A3: Treatment Effects (Inverse Probability Weighted)

	Family index: Education (1)	Family index: Cognition (2)	Family index: Socioemotional (3)	Family index: Labor outcomes (4)	Family index: SES/Well-being (5)
Treatment	0.186***	0.053	0.067	0.031	0.154***
	(0.067)	(0.066)	(0.064)	(0.054)	(0.055)
Observations	1,370	1,369	1,360	1,241	1,354
R-squared	0.123	0.151	0.044	0.105	0.110
F-statistic	3.848	8.007	2.934	4.250	6.308
Covariates	Yes	Yes	Yes	Yes	Yes
Control mean	0.039	0.018	-0.003	-0.001	-0.002

Notes. Estimated treatment effects, weighted by respondents' inverse probability of attriting (IPW). IPWs estimated through a probit regression of an attrition indicator on treatment status and the covariates listed in Table 1. The dependent variables in columns (1) to (5) are the family indices from Tables 3 to 8. Treatment captures effects for students who received any treatment, under either scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Heterogeneous and Labelling Effects (Inverse Probability Weighted)

	Heterogeneous Effects					Labelling Effects				
	Family index: Education (1)	Family index: Cognition (2)	Family index: Socioemotional (3)	Family index: Labor outcomes (4)	Family index: SES/Well-being (5)	Family index: Education (6)	Family index: Cognition (7)	Family index: Socioemotional (8)	Family index: Labor outcomes (9)	Family index: SES/Well-being (10)
Treatment (any)	0.142*	0.128	0.069	0.029	0.231***	0.127	0.235**	0.124	0.134	0.187**
	(0.077)	(0.079)	(0.078)	(0.063)	(0.069)	(0.103)	(0.105)	(0.112)	(0.089)	(0.089)
Treatment (poverty school)	0.090	-0.156*	-0.005	0.004	-0.158*	0.044	-0.193*	0.021	-0.083	-0.153
	(0.096)	(0.094)	(0.095)	(0.078)	(0.087)	(0.108)	(0.107)	(0.112)	(0.093)	(0.098)
Treatment (merit school and non-poor)						-0.011	-0.222	-0.076	-0.257**	0.083
						(0.128)	(0.138)	(0.154)	(0.130)	(0.120)
Treatment (poverty school and low scores)						0.179	-0.108	-0.200	0.049	0.076
						(0.111)	(0.140)	(0.157)	(0.145)	(0.159)
Observations	1,370	1,369	1,360	1,241	1,354	1,370	1,369	1,360	1,241	1,354
R-squared	0.124	0.154	0.044	0.105	0.113	0.126	0.155	0.045	0.108	0.113
Effect in Poverty Schools	0.232***	-0.028	0.064	0.033	0.073					
	(0.088)	(0.083)	(0.081)	(0.070)	(0.071)					
Effect in Poverty Schools, if high scores						0.171	0.041	0.144	0.052	0.035
						(0.105)	(0.104)	(0.109)	(0.090)	(0.088)
F-statistic	4.055	7.807	2.880	4.152	6.215	4.126	7.574	2.862	4.233	6.076
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.039	0.018	-0.003	-0.001	-0.002	0.060	0.033	-0.082	-0.061	-0.088

Notes. Estimated treatment effects, weighted by respondents' inverse probability of attriting (IPW). IPWs estimated through a probit regression of an attrition indicator on treatment status and the covariates listed in Table 1. The dependent variables in columns (1) to (10) are the family indices from Tables 3 to 5 and 7 to 8. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Appendix B Complementarities between cognitive and socioemotional outcomes

Our approach is based on two different conceptual models of the relationships between years of education (E), cognitive skills (C), and socioemotional skills (S). As a starting point, based on the evaluation three years after the program's inception (Barrera-Osorio and Filmer 2016), we know that treatment T_0 (at baseline, $t = 0$) increased years of education schooling for both merit- and poverty-based scholarships ($E_t = f(T_0; X_0, Z_0)$; $\frac{E_t}{\partial T_0} > 0$, for both types of scholarships). Furthermore, the evaluation showed a causal, positive effect of the intervention on cognitive skills for the merit-based scholarship only ($C_t^M = f(T_0^M; X_0, Z_0)$, $\frac{\partial C_t^M}{\partial T_0^M} > 0$); and zero effects for the poverty-based scholarship ($C_t^P = f(T_0^P; X_0, Z_0)$, $\frac{\partial C_t^P}{\partial T_0^P} = 0$), where M denotes merit-based treatment and P denotes poverty-based treatment.

The first conceptual relationship we explore is that between each type of skill—cognitive and socioemotional—and years of education:

$$C_t = g(E_t; X_0, Z_0)$$
$$S_t = g(E_t; X_0, Z_0)$$

, where X_0 are student characteristics and Z_0 are school inputs (at baseline). These equations state that the effect on either set of skills is a function of the years of education; i.e., exposure to more schooling will induce higher cognitive and socioemotional skills. Therefore, the first set of relationships we investigate are:

$$\frac{\partial C}{\partial T} = \frac{\partial C}{\partial E} * \frac{\partial E}{\partial T} > 0 \tag{4}$$

and

$$\frac{\partial S}{\partial T} = \frac{\partial S}{\partial E} * \frac{\partial E}{\partial T} > 0 \tag{5}$$

If schooling produces cognitive and socioemotional skills, both equations 4 and 5 are positive, independently of the type of treatment (merit or poverty).

In contrast, the second conceptual relationship is based on a modification of

this setup: for the merit-based scholarship we have an additional equation, relating cognitive skills and treatment:

$$C_t^M = f(T_0^M) \tag{6}$$

i.e., treatment induced higher cognitive skills only for the merit (M) treatment. The basic relationship of interest is between socioemotional skills and cognitive skills:

$$S_t^M = g(C_t^M, E_t; X_0, Z_0)$$

The second relationship we investigate is therefore:

$$\frac{\partial S_t^M}{\partial T_0^M} = \frac{\partial S_t^M}{\partial C_t^M} * \frac{\partial C_t^M}{\partial T_0^M} + \frac{\partial S_t^M}{\partial E_t} * \frac{\partial E_t}{\partial T_0^M} > 0 \tag{7}$$

i.e., that the effect of treatment on socioemotional skills is positive, and it depends on the effect of cognitive skills on socioemotional skills ($\frac{\partial S_t^M}{\partial C_t^M}$) and on the indirect effect of higher exposure to more schooling ($\frac{\partial S_t^M}{\partial E_t}$). If there is complementarity (or co-production) between cognitive and socioemotional skills (i.e., $\frac{\partial S_t^M}{\partial C_t^M} > 0$), then $\frac{\partial S_t^M}{\partial T_0^M} > 0$.

For the case of the poverty-based scholarship (P), the corresponding expression is:

$$\frac{\partial S_t^P}{\partial T_0^P} = \frac{\partial S_t^P}{\partial E_t} * \frac{\partial E_t}{\partial T_0^P} \tag{8}$$

since

$$\frac{\partial C_t^P}{\partial T_0^P} = 0$$

. There are three main relevant cases for Equations 7 and 8. If exposure to school in-and-of itself produces socioemotional skills, both Equation 7 and 8 are positive. If exposure to schooling does not produce socioemotional skills, Equation 8 is equal to zero. Finally, under complementarities between cognitive and socioemotional skills (e.g. if cognitive skills help in the acquisition of socioemotional skills, or if they are co-produced), then Equation 7 is positive, independent of the relationship between socioemotional skills and exposure to school.

Online Appendix

Long-Term Impacts of Alternative Approaches to Increase Schooling:
Experimental Evidence from a Scholarship Program in Cambodia

Felipe Barrera-Osorio   Andreas de Barros   Deon Filmer
Harvard University   Harvard University   The World Bank

Table OA1: Heterogeneous and Labelling Effects on Education Outcomes

	Heterogeneous Effects			Labelling Effects		
	Highest grade completed	Completed primary	Received any formal education in 2011-2017	Highest grade completed	Completed primary	Received any formal education in 2011-2017
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (any)	0.253*	0.055	0.042	0.265	0.036	0.038
	(0.130)	(0.040)	(0.031)	(0.173)	(0.054)	(0.042)
Treatment (poverty school)	-0.026	0.052	0.055	-0.034	0.054	0.028
	(0.149)	(0.048)	(0.039)	(0.160)	(0.055)	(0.043)
Treatment (merit school and non-poor)				-0.027	0.035	-0.017
				(0.220)	(0.067)	(0.051)
Treatment (poverty school and low scores)				0.002	0.036	0.097*
				(0.190)	(0.067)	(0.055)
Observations	1,370	1,370	1,370	1,370	1,370	1,370
R-squared	0.110	0.107	0.089	0.110	0.107	0.091
Effect in Poverty Schools	0.227	0.107***	0.097**			
	(0.142)	(0.041)	(0.038)			
Effect in Poverty Schools, if high scores				0.231	0.090*	0.066
				(0.175)	(0.051)	(0.044)
F-statistic	3.104	3.597	3.709	2.988	3.457	3.769
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	5.478	0.573	0.745	5.530	0.595	0.744

Notes. Estimated treatment effects. Dependent variables as in Table 3. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table OA2: Heterogeneous and Labelling Effects on Cognitive Outcomes

	Heterogeneous Effects				Labelling Effects			
	Math	Raven's	Forward Digit Span	Picture Recognition Vocabulary Test	Math	Raven's	Forward Digit Span	Picture Recognition Vocabulary Test
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (any)	0.089	0.139**	0.079	0.066	0.230**	0.192**	0.096	0.178
	(0.083)	(0.070)	(0.065)	(0.084)	(0.107)	(0.095)	(0.089)	(0.115)
Treatment (poverty school)	-0.057	-0.102	-0.232***	-0.035	-0.139	-0.103	-0.214**	-0.053
	(0.092)	(0.089)	(0.082)	(0.101)	(0.105)	(0.098)	(0.096)	(0.114)
Treatment (merit school and non-poor)					-0.315**	-0.092	-0.012	-0.207
					(0.138)	(0.126)	(0.148)	(0.138)
Treatment (poverty school and low scores)					-0.027	-0.106	-0.096	-0.178
					(0.140)	(0.149)	(0.146)	(0.139)
Observations	1,370	1,370	1,369	1,370	1,370	1,370	1,369	1,370
R-squared	0.111	0.122	0.054	0.214	0.114	0.123	0.054	0.216
Effect in Poverty Schools	0.032	0.038	-0.153**	0.031				
	(0.072)	(0.077)	(0.075)	(0.094)				
Effect in Poverty Schools, if high scores					0.091	0.088	-0.118	0.125
					(0.095)	(0.093)	(0.096)	(0.118)
F-statistic	6.441	7.981	2.421	9.493	6.128	7.803	2.341	9.093
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.023	0.007	0.010	0.031	0.005	0.038	0.043	0.009

Notes. Estimated treatment effects. Dependent variables as in Table 4. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table OA3: Heterogeneous and Labelling Effects on Emotional and Behavioral Difficulties

	Heterogeneous Effects			Labelling Effects		
	Prosocial (1)	Internalizing (2)	Externalizing (3)	Prosocial (4)	Internalizing (5)	Externalizing (6)
Treatment (any)	-0.011	-0.109*	-0.026	0.137	-0.036	-0.007
	(0.066)	(0.064)	(0.076)	(0.096)	(0.099)	(0.100)
Treatment (poverty school)	0.022	0.054	0.039	0.025	0.033	-0.024
	(0.080)	(0.083)	(0.091)	(0.092)	(0.098)	(0.102)
Treatment (merit school and non-poor)				-0.253	-0.144	-0.089
				(0.157)	(0.147)	(0.134)
Treatment (poverty school and low scores)				-0.326**	-0.087	0.168
				(0.144)	(0.164)	(0.158)
Observations	1,367	1,368	1,368	1,367	1,368	1,368
R-squared	0.040	0.075	0.041	0.045	0.076	0.042
Effect in Poverty Schools	0.010	-0.055	0.013			
	(0.075)	(0.079)	(0.071)			
Effect in Poverty Schools, if high scores				0.162*	-0.003	-0.031
				(0.093)	(0.106)	(0.101)
F-statistic	1.882	4.036	1.594	2.062	4.070	1.581
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	-0.014	-0.003	-0.004	-0.137	0.010	0.007

Notes. Estimated treatment effects. Dependent variables as in Table 5. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table OA4: Heterogeneous and Labelling Effects on Personality Traits

	Heterogeneous Effects					Labelling Effects				
	Openness (1)	Conscientiousness (2)	Extraversion (3)	Agreeableness (4)	Neuroticism (5)	Openness (6)	Conscientiousness (7)	Extraversion (8)	Agreeableness (9)	Neuroticism (10)
Treatment (any)	0.029	-0.072	0.035	-0.054	-0.066	0.132	-0.018	0.047	0.013	-0.045
	(0.067)	(0.065)	(0.076)	(0.068)	(0.073)	(0.103)	(0.087)	(0.108)	(0.090)	(0.102)
Treatment (poverty school)	0.002	0.069	-0.148	-0.007	-0.092	-0.035	0.110	-0.130	-0.044	-0.111
	(0.079)	(0.075)	(0.090)	(0.077)	(0.083)	(0.089)	(0.086)	(0.109)	(0.090)	(0.096)
Treatment (merit school and non-poor)						-0.211	-0.056	-0.005	-0.147	-0.052
						(0.157)	(0.145)	(0.147)	(0.136)	(0.138)
Treatment (poverty school and low scores)						-0.008	-0.250*	-0.083	-0.022	0.019
						(0.163)	(0.150)	(0.168)	(0.161)	(0.182)
Observations	1,368	1,369	1,368	1,368	1,370	1,368	1,369	1,368	1,368	1,370
R-squared	0.036	0.034	0.054	0.036	0.024	0.038	0.036	0.054	0.036	0.024
Effect in Poverty Schools	0.031	-0.003	-0.113	-0.061	-0.158**					
Effect in Poverty Schools, if high scores	0.068	0.071	0.078	0.071	0.071	0.097	0.092	-0.083	-0.031	-0.156
	(0.091)	(0.086)	(0.109)	(0.085)	(0.102)	(0.091)	(0.086)	(0.109)	(0.085)	(0.102)
F-statistic	2.048	2.481	2.031	1.541	1.366	2.159	2.346	1.948	1.503	1.323
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.009	0.001	0.016	0.000	0.022	-0.044	-0.030	-0.032	0.014	0.029

Notes. Estimated treatment effects. Dependent variables as in Table 5. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme, but would have also qualified under the merit-based scheme. Effect in Poverty Schools, if high scores captures effects for students under the merit-based scheme, but would have also qualified under the merit-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table OA5: Heterogeneous and Labelling Effects on Labor Outcomes

| | Heterogeneous Effects | | | | | | Labelling Effects | | | | | |
	Currently working	Age started working	Any training since 2011	Cog. demands of main work (1/0)	Yearly earnings (inv. hyperbolic sine, USD)	Daily res. wage (inv. hyperbolic sine, USD)	Currently working	Age started working	Any training since 2011	Cog. demands of main work (1/0)	Yearly earnings (inv. hyperbolic sine, USD)	Daily res. wage (inv. hyperbolic sine, USD)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment (any)	0.028	0.047	-0.024	-0.029	-0.214	0.068	0.016	0.304	-0.019	-0.011	-0.469*	0.097*
	(0.020)	(0.214)	(0.040)	(0.025)	(0.188)	(0.042)	(0.029)	(0.273)	(0.050)	(0.034)	(0.277)	(0.051)
Treatment (poverty school)	-0.003	0.263	-0.012	0.000	-0.004	-0.050	0.007	0.176	-0.004	0.000	0.107	-0.090
	(0.023)	(0.256)	(0.046)	(0.030)	(0.206)	(0.048)	(0.028)	(0.271)	(0.052)	(0.034)	(0.301)	(0.055)
Treatment (merit school and non-poor)							0.029	-0.315	-0.002	-0.055	0.537	-0.085
							(0.038)	(0.410)	(0.066)	(0.050)	(0.388)	(0.075)
Treatment (poverty school and low scores)							-0.009	-0.259	-0.039	0.059	0.176	0.069
							(0.042)	(0.484)	(0.071)	(0.052)	(0.412)	(0.093)
Observations	1,300	1,327	1,370	1,299	1,334	1,353	1,300	1,327	1,370	1,299	1,334	1,353
R-squared	0.062	0.041	0.092	0.056	0.072	0.072	0.063	0.042	0.092	0.059		0.073
Effect in Poverty Schools	0.025	0.310	-0.036	0.001	-0.217	0.018						
	(0.018)	(0.222)	(0.036)	(0.027)	(0.227)	(0.047)						
Effect in Poverty Schools, if high scores							0.023	0.480*	-0.022	-0.010	-0.362	0.008
							(0.029)	(0.281)	(0.047)	(0.034)	(0.261)	(0.057)
F-statistic	2.465	1.895	4.363	2.341	2.445	2.058	2.423	1.941	4.230	2.410	2.372	1.993
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.919	12.637	0.588	0.159	7.460	2.222	0.907	12.565	0.567	0.155	7.603	2.213

*Notes: Estimated treatment effects. Dependent variables as in Table 7. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the merit-based scheme. Effect in Poverty Schools captures effects for students who received the treatment under the poverty-based scheme, if high scores captures effects for students who would have also qualified under the treatment under the poverty-based scheme, but would have also qualified under the merit-based scheme. All regression control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.*

Table OA6: Heterogeneous and Labelling Effects on Well-being Outcomes

	Heterogeneous Effects						Labelling Effects					
	SES Ladder (village)	SES Index (IRT)	Life Satisfaction	Quality of Health	Quality of Life	Health Issue Index (GHQ)	SES Ladder (village)	SES Index (IRT)	Life Satisfaction	Quality of Health	Quality of Life	Health Issue Index (GHQ)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment (any)	0.210***	0.159**	0.108	0.201***	0.076	-0.068	0.212**	0.180*	0.080	0.210**	0.061	0.025
	(0.077)	(0.076)	(0.067)	(0.066)	(0.062)	(0.069)	(0.101)	(0.093)	(0.092)	(0.098)	(0.080)	(0.090)
Treatment (poverty school)	-0.072	-0.204**	-0.059	-0.149*	0.031	0.047	-0.136	-0.199**	-0.043	-0.193*	0.034	-0.065
	(0.085)	(0.088)	(0.076)	(0.086)	(0.078)	(0.086)	(0.095)	(0.083)	(0.087)	(0.099)	(0.091)	(0.097)
Treatment (north school and non-poor)							-0.060	-0.033	0.002	-0.056	0.028	-0.207
							(0.134)	(0.120)	(0.124)	(0.142)	(0.127)	(0.126)
Treatment (poverty school and low scores)							0.209	-0.064	0.005	0.126	0.019	-0.026
							(0.154)	(0.155)	(0.159)	(0.160)	(0.159)	(0.167)
Observations	1,370	1,370	1,369	1,370	1,370	1,355	1,370	1,370	1,369	1,370	1,370	1,355
R-squared	0.073	0.221	0.075	0.051	0.052	0.067	0.074	0.221	0.076	0.052	0.052	0.059
Effect in Poverty Schools	0.138**	-0.046	0.048	0.052	0.107	-0.021						
	(0.070)	(0.073)	(0.062)	(0.071)	(0.070)	(0.075)						
Effect in Poverty Schools, if high scores							0.076	-0.019	0.037	0.017	0.096	0.020
							(0.093)	(0.087)	(0.086)	(0.090)	(0.085)	(0.097)
F-statistic	3.880	10.670	3.551	2.353	2.691	4.063	3.927	10.405	3.409	2.275	2.597	4.119
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	-0.014	0.002	-0.020	0.008	0.012	-0.011	-0.065	-0.088	-0.088	0.003	-0.052	-0.018

Notes: Estimated treatment effects. Dependent variables as in Table 8. In the panel to the left (heterogeneous effects), treatment (any) captures effects for students who received the treatment, under the poverty-based scheme. Effect in Poverty Schools captures effects for students who received the treatment, under the poverty-based scheme. In the panel to the right (labelling effects), treatment (any) captures effects for students who received the treatment under the north-based scheme, but would have also qualified under the poverty-based scheme. Effect in Poverty Schools, if high scores captures effects for students who received the treatment under the poverty-based scheme, but would have also qualified under the north-based scheme. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table OA7: Heterogeneous Treatment Effects by Gender: Education Outcomes

	Highest grade completed	Completed primary	Received any formal education in 2011-2017
	(1)	(2)	(3)
Treatment	0.199	0.068	0.053
	(0.129)	(0.043)	(0.035)
Female and treatment	0.069	0.016	0.028
	(0.175)	(0.061)	(0.044)
Female	-0.120	0.014	-0.055
	(0.178)	(0.054)	(0.043)
Observations	1,337	1,337	1,337
R-squared	0.114	0.108	0.088
Effect for females	0.268*	0.084*	0.081**
	(0.155)	(0.045)	(0.037)
F-statistic	3.243	3.416	3.513
Covariates	Yes	Yes	Yes
Control mean (females)	5.385	0.565	0.723

Notes. Estimated treatment effects. Dependent variables as in Table 3. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table OA8: Heterogeneous Treatment Effects by Gender: Cognitive Outcomes

	Math	Raven's	Forward Digit Span	Picture Recognition Vocabulary Test
	(1)	(2)	(3)	(4)
Treatment	0.088	0.193**	0.084	0.211**
	(0.089)	(0.086)	(0.085)	(0.096)
Female and treatment	-0.057	-0.205*	-0.211*	-0.307***
	(0.111)	(0.118)	(0.111)	(0.114)
Female	-0.182*	-0.261**	-0.003	-0.174
	(0.095)	(0.107)	(0.108)	(0.108)
Observations	1,337	1,337	1,336	1,337
R-squared	0.112	0.124	0.050	0.219
Effect for females	0.031	-0.012	-0.127*	-0.096
	(0.081)	(0.081)	(0.076)	(0.089)
F-statistic	6.317	8.258	2.194	11.646
Covariates	Yes	Yes	Yes	Yes
Control mean (females)	-0.075	-0.136	0.006	-0.046

Notes. Estimated treatment effects. Dependent variables as in Table 4. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table OA9: Heterogeneous Treatment Effects by Gender: Socioemotional Outcomes

	SDQ			Big 5				
	Prosocial (1)	Internalizing (2)	Externalizing (3)	Openness (4)	Conscientiousness (5)	Extraversion (6)	Agreeableness (7)	Neuroticism (8)
Treatment	0.028	-0.134	-0.020	-0.019	-0.040	0.068	-0.051	-0.102
	(0.077)	(0.087)	(0.088)	(0.084)	(0.085)	(0.086)	(0.088)	(0.082)
Female and treatment	-0.045	0.092	0.023	0.100	0.014	-0.178	-0.003	-0.000
	(0.103)	(0.115)	(0.120)	(0.114)	(0.116)	(0.120)	(0.116)	(0.114)
Female	0.098	0.234**	0.231**	-0.074	-0.134	0.278***	-0.043	-0.056
	(0.098)	(0.101)	(0.098)	(0.093)	(0.112)	(0.102)	(0.119)	(0.099)
Observations	1,334	1,336	1,335	1,335	1,336	1,335	1,335	1,337
R-squared	0.043	0.077	0.042	0.039	0.036	0.054	0.037	0.024
Effect for females	-0.017	-0.042	0.003	0.082	-0.026	-0.109	-0.053	-0.102
	(0.079)	(0.077)	(0.080)	(0.075)	(0.076)	(0.086)	(0.076)	(0.082)
F-statistic	2.081	4.568	1.658	2.054	2.261	1.874	1.793	1.606
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean	0.051	0.159	0.084	-0.038	-0.047	0.132	0.040	0.039

Notes. Estimated treatment effects. Dependent variables as in Table 5. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table OA10: Heterogeneous Treatment Effects by Gender: Labor Outcomes

	Currently working	Age started working	Any training since 2011	Cog. demands of main work (1/0)	Yearly earnings (inv. hyperbolic sine, USD)	Daily res. wage (inv. hyperbolic sine, USD)
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.035**	0.271	-0.000	-0.007	-0.160	0.072
	(0.017)	(0.265)	(0.042)	(0.031)	(0.224)	(0.058)
Female and treatment	-0.018	-0.218	-0.053	-0.024	-0.131	-0.053
	(0.026)	(0.357)	(0.054)	(0.042)	(0.309)	(0.074)
Female	-0.045*	-0.334	-0.145***	0.022	-0.728**	-0.091
	(0.024)	(0.321)	(0.045)	(0.039)	(0.284)	(0.064)
Observations	1,267	1,295	1,337	1,267	1,301	1,320
R-squared	0.068	0.039	0.092	0.055		0.068
Effect for females	0.017	0.054	-0.053	-0.030	-0.291	0.019
	(0.022)	(0.241)	(0.040)	(0.030)	(0.223)	(0.047)
F-statistic	2.586	1.916	4.690	2.061	2.404	2.077
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Control mean (females)	0.891	12.449	0.511	0.167	7.265	2.178

Notes. Estimated treatment effects. Dependent variables as in Table 7. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** p<0.01, ** p<0.05, * p<0.1.

Table OA11: Heterogeneous Treatment Effects by Gender: Well-being Outcomes

	SES Ladder (village) (1)	SES Index (IRT) (2)	Life Satisfaction (3)	Quality of Health (4)	Quality of Life (5)	Health Issue Index (GHQ) (6)
Treatment	0.275***	0.176**	-0.007	0.256***	0.087	-0.006
	(0.078)	(0.080)	(0.079)	(0.077)	(0.079)	(0.080)
Female and treatment	-0.194*	-0.207*	0.153	-0.226*	0.002	-0.066
	(0.116)	(0.106)	(0.118)	(0.121)	(0.108)	(0.109)
Female	0.029	0.079	-0.240**	-0.063	-0.003	0.237**
	(0.100)	(0.084)	(0.115)	(0.095)	(0.097)	(0.099)
Observations	1,337	1,337	1,336	1,337	1,337	1,322
R-squared	0.077	0.221	0.077	0.053	0.052	0.054
Effect for females	0.080	-0.030	0.146*	0.030	0.089	-0.072
	(0.088)	(0.086)	(0.079)	(0.083)	(0.074)	(0.079)
F-statistic	4.216	10.516	3.559	2.378	2.835	3.255
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Control mean (females)	-0.016	-0.025	-0.172	-0.043	-0.038	0.122

Notes. Estimated treatment effects. Dependent variables as in Table 8. Treatment captures effects for male students who received any treatment, under either scheme. Effect for females captures the respective effect for female students. All regressions control for province fixed effects, baseline test score, baseline poverty score, individual-level socio-economic variables from baseline, 6 school-level (EMIS) variables and 5 census village-level variables, measured at baseline. Standard errors are in parentheses (clustered at the school level). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.