The Impact of Girls' Education Support Program on Human Capital Development: Evidence from a Randomized Evaluation in Malawian Secondary Schools *

Booyuel Kim†

September 26, 2014

Abstract

This paper evaluates a randomized controlled trial of the girls' education support program on human capital development of 3,997 female students (9^{th} - 11^{th}) across 124 classrooms in 33 public secondary schools in Malawi. We find that female students treated with one-year tuition support and monthly cash stipends are more likely to attend school and have better test scores. We also find that cognitive ability in the treatment group increases by 0.215 standard deviations, and those treated also display higher aspirations for educational achievement. Moreover, there is a significant improvement in time preference (increased patience). (JEL: C93, I20, O15)

*I wish to thank to Cristian Pop-Eleches, Hyuncheol Bryant Kim, Miguel Urquiola, Douglas Almond, Matthew Neidell and Prabhjot Singh for helpful feedback. I also thank Young-sim Baek, Susie Kim, Sang-eun Park, Joshua Kim, Jinho Kim, Taeksoo Kim, Hanyoun So and all members of 'Project Malawi' team for their cooperation, excellent field work, and data management. This project is supported by funding and administrative support from Korea International Cooperation Agency (KOICA), Africa Future Foundation (AFF), and Daeyang Luke Hospital. All views expressed are mine, and all errors are my own.

†bk2267@columbia.edu, Earth Institute, Columbia University, New York, New York 10115.

1 Introduction

Education has been considered one of the most important determinants of human capital development. It is argued that improved education (mostly measured by test scores of reading comprehension and mathematics) is correlated with increased wages later in life at the micro-level (Blau and Kahn 2005) and may lead to more economic growth of a country at the macro-level (Hanushek and Woessmann 2012). However, previous studies assessing causal impacts of various educational interventions on test scores are limited to explaining whether these test score gains translate into improvements in adult economic and social outcomes.

One of the few exceptions is Chetty et al. (2011) who evaluate the impact of Project STAR (Student/Teacher Achievement Ratio) in Tennessee on college attendance, earnings, retirement savings, home ownership, and marriage 20 years after the implementation of the program. They present long-term impacts of smaller student-to-teacher ratio at $K - 3^{rd}$ grades on outcomes in adulthood. Barnett (1992) also presents similar results showing that the Perry pre-school program in Michigan has produced long-term impacts on adult outcomes. Interestingly, they both report that the impacts on test scores were ephemeral and fade out rapidly but nonetheless educational innovations (small classroom or pre-school education) had significant long-term impacts. For example, Krueger and Whitmore (2001) shows that the impacts of class size from the Tennessee STAR experiment on test scores become statistically insignificant by grade 8 but then they reemerge for adult outcomes (Chetty et al. 2011). However, they do not shed light on which specific factors other than test scores have affected long-term outcomes.

Recent studies emphasize the importance of cognitive and non-cognitive abilities as well as time and risk preferences as underlying mechanisms to connect educational interventions to test scores and later adulthood outcomes (Macours et al. 2012; Wydick et al. 2013; Glewwe et al. 2013; Sutter et al. 2013). We contribute to this literature by providing a unified evaluation of several outcomes, including the first analysis of causal impacts on preferences. Our analysis attempts to capture the effect that the educational intervention has on a number of intermediate outcome variables (such as cognitive and non-cognitive skills, and time preferences) that could be potential mechanisms explaining outcomes during adulthood. The ideal way to identify possible underlying mechanism of educational interventions for outcomes in adulthood is a randomized controlled trial targeted towards experimentally changing each potential mechanism separately. While the present study was not designed to do this, it is nevertheless important to capture and describe the impact of the educational intervention on these potential mechanisms in the short run.

In this paper, we use experimental data from the girls' education support program in Malawi. In 2012, 124 classrooms of grade 9, 10, and 11 across 33 public secondary schools in Lilongwe rural areas participated in the girls' education support program and 62 classrooms were randomly selected as the treatment group while the remaining 62 classrooms served as the control. All female students in treatment classrooms received one-year tuition directly deposited to the school account as well as monthly stipends distributed to students. By exploiting this experimental design, we evaluate the causal impacts of the education support program on schooling outcomes (dropout and absence), test scores, cognitive/non-cognitive abilities, and time preferences.

Starting with schooling outcomes, we find that although dropout rates (both self-reported and school-reported) declined, it was statistically significant only in 9^{th} grade (starting grade of a secondary school) students whereas it is insignificant in 10^{th} and 11^{th} grade students. This is consistent with recent work by Son (2013) who finds that negative income shocks (unemployment, crop loss, drought, and Asian financial crises) in Indonesia affect school enrollment differentially across different grade levels and this impact is strongly mitigated for students who enter the final grades of junior or senior high school.[1] The probability of remaining at the original school in the treatment group (taking both dropouts and transfer students into account) increased by 8.7 percentage points at the 99% confidence level. School absence also declined among treated students on average by 5 days per year. When we turn to examine the Malawi School Certificate Exam (MSCE) and Junior Certificate Exam (JCE) for baseline 11^{th} grade and 9^{th} grade students, respectively, program impacts on MSCE were small and statistically insignificant in the treatment group. But we find evidence that the scholarship program increased the probability to take JCE for grade 9 students in the treatment group by 17.3 percentage points. The probability to pass JCE also raised by 15.1 percentage points at the 99% confidence level and overall exam scores improved by 0.241 SD when we control for baseline characteristics. However, the improvements in JCE performance need to be interpreted cautiously because students in the treatment classrooms took the JCE more than the

[1] This is so called sheepskin effect. A sheepskin effect exits when the wage return to an additional year of schooling is higher if that year allows a student to complete a school level. The origin of the term relates to the fact that diplomas were once printed on sheepskin.

control group, which may cause selection bias in the first stage.

Our main findings are that cognitive and non-cognitive abilities of treatment group students improved. Cognitive ability in the treatment increased by 0.215 SD at the 99% confidence level and treated students take education more importantly with higher aspirations for educational achievement. We also find significant improvements in time preferences and annual discount rate declined by 36.7 percentage points on average among students in the treatment classrooms at the 95% significant level. This causal impact of the girls education support program on time preferences complements the previous studies on time and risk preferences, which only present correlations between preferences and interested outcomes such as demographics, education, health, and labor market outcomes.

The paper is organized as follows. In Section 1.2, we briefly review related literature on education support program. In Section 1.3, we present the experimental design and address potential biases to the validity of the experiment. In Section 1.4, the empirical strategy based on simple OLS models is presented, followed by the results and analysis on schooling outcomes, test scores, cognitive/non-cognitive abilities, and time preferences in Section 1.5. Finally, Sections 1.6 concludes.

2 A Brief Literature Review on Education Support Programs

This section provides a brief overview of the related literature, limiting the scope to only financial incentives on students from field experiments. Randomized evaluations on financial incentives for education improvements have been concentrated in developing countries. Kremer, Miguel, and Thornton (2009) evaluate the effect of a merit-based girls' scholarship program in rural Kenya. The top 15% of 6th grade female students in the program districts received the scholarship (US $20 per year) for two years. They find that the merit scholarship improved average test scores by 0.19 SD.

Colombia's school voucher program (PACES) with a merit-scholarship component provided nearly 125,000 students from poor neighborhoods with vouchers (worth US $190) between 1991 and 1997 that covered approximately half the cost of private secondary school. Angrist et al. (2002) find that treated students were more likely to attend private school with improved test scores. 7 years after the voucher program, lottery winners were more likely to graduate from high school and scored higher on college entrance exams (Angrist, Bettinger, and Kremer, 2006).

The pioneering conditional cash transfer program in Mexico, PROGRESA provided monthly cash grants (US $55 on average) to poor family, with conditionality that their children should attend school at least 85% of the time. Schultz (2004) reported increase in school enrollment and Behrman, Sengupta, and Todd (2005) showed that participation in the PROGRESA program is associated with not only improved school enrollment but also less grade repetition, lower dropout rates, and higher school reentry rates among dropouts.

The work similar to this chapter is that of Baird, McIntosh, and Özler (2011) who used an experimental design to estimate the effects of conditional (and unconditional) cash transfers on schooling outcomes, test scores, marriage, and pregnancy for never-married girls, aged 13-22 in Malawi. They find that school enrollment and attendance as well as test scores were significantly higher among treated in-school girls but there was no impact on marriage and pregnancy among them, which is consistent with findings in this paper. A recent cash transfer program in rural Morocco shows that an unconditional cash transfer (US $80 - $130 per year) made to households of primary school age children had a very large impact on school participation despite the fact that the transfer was not conditional on regular school attendance (Benhassine et al., 2013).

In general, various financial incentives such as merit scholarship, school vouchers, and conditional cash transfers on students' education have been proved effective to improve schooling outcomes and test scores.

3 Background, Experimental Design and Data

3.1 Background: Education in Malawi

Malawi is a small landlocked country in Sub-Saharan Africa and basic education in Malawi consists of eight years of primary education (Standard 1 through 8) followed by four years of secondary education (Form 1 through 4). Unlike the universal primary education policy adopted in 1994, secondary school admission depends on performance on the Primary School Leaving Certificate Examination (PSLCE). Secondary students have to pay school tuition and fees and these costs averaged approximately US$ 21 (Malawi Kwacha 3,500) per semester.[2] Students have to pay tuition and fees each semester and if they dont submit the payment until the first couple of weeks of each semester, then they would be unable to enroll the school and be dropped out from the school. 10^{th} grade students (Form 2) must pass the Junior Certificate Examination (JCE) at the end of their second year in secondary school in order to move on to the next grade and 12^{th} grade students (Form 4) take the Malawi School Certificate Examination (MSCE) before they graduate from secondary school. Students must pay the fee (MK 300 for JCE and MK 420 for MSCE) to take the exam and if 10^{th} grade students fail to pass JCE, then they have to repeat 10^{th} grade again either in the original school or a transferred school. If 12^{th} grade students fail to pass MSCE, then they graduate from a secondary school without the certificate and would be unable to apply for tertiary education until they pass MSCE.

3.2 Experimental Design

We implemented one-year education support program to 3,997 girls (9^{th} - 11^{th} grades) at 33 public schools located around rural Lilongwe, Malawi.[3] This program was part of a broader HIV/AIDS prevention program which includes HIV/AIDS education, male circumcision, and the girls' education support program being implemented by Daeyang Luke Hospital in Lilongwe, Malawi.[4] Table 1 shows the experimental designs for the three HIV/AIDS prevention programs, which were independently implemented. For the girls' education support program, we first stratified 124 classrooms by grade and randomly assigned 62 classrooms the treatment status. All girl students in the treatment classrooms received one-year school tuition (three semesters) and monthly cash stipends (three times per semester). School tuition and fees per semester on average was 3,500 Malawi kwacha and were directly deposited to each schools account and monthly cash stipends of 300 Malawi kwacha

[2]Secondary schools in Malawi run 3 semesters per year. First semester starts in September to December. Second semester is January to April and third semester is April to July. One US dollar was worth 165 Malawi kwachas (MK) in April 2012 (http://www.oanda.com/currency/converter).

[3]The target population of 9^{th} - 11^{th} grade female students was selected for a variety of reasons. As this study was designed as part of a larger HIV/AIDS prevention program, we focused on females because the HIV prevalence rate among boys of secondary schooling age is negligible. The grade range was selected so that the study population had a reasonable chance of being or becoming sexually active during the study period. Finally, a decision was made to not make any offers to 12^{th} graders who were about to graduate before the one-year follow-up survey because of budget constraint and logistical challenge for tracking after graduation.

[4]The partner hospital has four catchment districts: Chimutu, Chitukula, Tsabango, and Kalumba in rural Lilongwe areas. Appendix 1 shows the map for these four districts. We invited all 33 public schools (excluding private boarding schools) in these districts to participate in the girls' education support program.

were distributed to treated students, which is equivalent to in total around US$ 70 per year.[5] These were substantial considering that Malawian GDP per capita (2013 est.) is US$ 224 and the minimum wage per month in rural areas is around US$17 (MK 2,742).[6] School tuition and fees for the beneficiary students were transferred to school account in the beginning of each semester while enumerators visited all schools every month on announced dates to directly distribute the monthly cash stipends to the treatment group. This program has a weak conditionality on school enrollment (not school participation) and beneficiary students must be enrolled in baseline school at the time of the transfers. The intervention was immediately discontinued for transfer and dropout students. The fact that the lottery was held with all 33 participating school headmasters under the supervision of the division education officer ensured that the process was transparent and helped the participating schools view the offers as fair.

[5] We started the girls education support program in the third semester of academic year 2011-2012 (April, 2012) and continued the intervention in the first and second semesters of academic year 2012-2013. We increased the amount of monthly stipend from 300 Kwacha to 500 Kwacha in the second semester of academic year 2012-13 due to the huge depreciation of Malawi Kwacha. Early in 2012, the exchange rate between US Dollar and Malawi Kwacha was 165/$. However, the value of Kwacha kept being depreciated after more than 50% currency devaluation on May 7^{th}, 2012 and the exchange rate in early 2013 were 350/$. In total, a treated student in the scholarship program received 13,900 Kwacha (3,500 Kwacha * 3 semesters + 300 Kwacha * 6 times + 500 Kwacha * 3 times), which is equivalent to around US $70 per year. Kremer, Miguel, and Thornton (2009) provided 6^{th} grade girls in Kenya with scholarship of US $20 per year and Baird et al. (2010) provided conditional cash transfers of US $120 per year to 13-22 year-old females in Malawi including current schoolgirls and recent dropouts.

[6] 81 percent of the employed persons in Malawi are self-employed subsistence farmer and only 8 percent of the employed persons are salaried workers (NSO 2011). The most widely used wage rates in Malawi are minimum wages. Although there is no information about how many workers are paid minimum wages, most firms seem to use them for low-skilled labor, which comprise the vast majority of the wage earners (ILO 2010).

Table 1: Experimental Designs

1) HIV/AIDS Education

	Group	Assignment	Classrooms	Students
100% Treatment	G1	Treatment	41	2,480
50% Treatment	G2	Treatment	41	1,303
	G3	No treatment		1,263
No Treatment	G4	No treatment (Control)	42	2,925
Total			124	7,971

2) Male Circumcision

	Group	Assignment	Classrooms	Students
100% Treatment	G1	Treatment	41	1,293
50% Treatment	G2	Treatment	41	679
	G3	No treatment		679
No Treatment	G4	No treatment (Control)	42	1,323
Total			124	3,974

3) Girls' Education Support

	Group	Assignment	Classrooms	Students
100% Treatment	G1	Treatment	62	2,102
No Treatment	G2	No treatment (Control)	62	1,895
Total			124	3,997

Notes: For HIV/AIDS education and Male circumcision interventions, the randomization was done in two stages. First, classrooms for each grade across 33 schools were randomly assigned to 100% treatment, 50% treatment and no treatment. Then, within 50% treatment, only half of the students were randomly assigned to treatment.

The validity of the experimental design which allows us to assess the causal inferences rests on two assumptions: successful randomization of students into treatment and control classrooms and no differences in attrition. We evaluate whether students were randomly assigned to the treatment and control groups. If students were randomly assigned, then the treatment indicator variable should not predict any predetermined baseline characteristics of the students. Table II shows baseline statistics and randomization balance for the all three HIV/AIDS prevention programs. The age of students on average is 16.2 years old and 5.4% of the sample are orphans without both parents. 19.8% and 8.2% of the sample reported that their fathers and mothers graduated from a 2-year college or 4-year university, respectively. Moreover, 45.6%, 32.7%, and 25.1% of the sample reported that their house has electricity, refrigerator and car at home, respectively (not shown).[7]

[7]Malawi DHS 2010 reported that only 9.1%, 4.3%, and 2.1% of the population have electricity, refrigerator, and car, respectively. Although studenst responses in household assets may be exaggerated, the huge differences between the sample and Malawi DHS 2010 can be understood after taking into account that our sample represents the family who is able to send their daughters to a secondary school.

Overall, our sample exhibits higher socioeconomic characteristics than the Malawi population as a whole. As shown in columns (2) - (8) of Table 2, none of the demographic characteristics predict the likelihood that a student is assigned to either one of three HIV/AIDS prevention interventions. F tests for the joint significance of all the predetermined demographic variables on HIV/AIDS education, male circumcision, and girls scholarship are insignificant ($p = 0.86$, 0.77, and 0.72) in columns (2), (3), and (6) respectively, showing that randomization for each intervention was well balanced across predetermined baseline characteristics. Column (7) specifically presents the randomization balance among the sub-sample of eligible girls and most of the baseline characteristics were well balanced between treatment and control classrooms except fathers education. However, the p-value of joint F test from a regression of full set of baseline controls on treatment assignment is 0.325 and it does not reject that all baseline coefficients are jointly equal to zero in the subsample of eligible girls only, either.

Table 2: Baseline statistics and Randomization Balance

Dependent Variable:	Avg. (s.d) (1)	HIV/AIDS Education (2)	Male Circumcision (3)	Male Circumcision (eligible boys) (4)	Male Circumcision (ineligible girls) (5)	Education Support (6)	Education Support (eligible girls) (7)	Education Support (ineligible boys) (8)
Age (year)	16.16	-0.008	0.010	-0.006	-0.007	0.000	0.003	0.001
	(1.856)	(0.006)	(0.008)	(0.007)	(0.011)	(0.008)	(0.011)	(0.008)
Orphan	0.054	-0.009	-0.016	-0.027	-0.006	-0.010	-0.037	0.013
	(0.227)	(0.025)	(0.024)	(0.037)	(0.029)	(0.024)	(0.032)	(0.035)
Father's tertiary education	0.198	0.009	0.010	-0.011	0.020	0.014	0.047**	-0.021
	(0.399)	(0.019)	(0.017)	(0.027)	(0.022)	(0.019)	(0.022)	(0.029)
Mother's tertiary education	0.082	-0.020	0.009	-0.014	0.038	0.033	-0.024	0.104***
	(0.274)	(0.027)	(0.025)	(0.038)	(0.030)	(0.024)	(0.030)	(0.037)
Father's white-collar job	0.256	-0.006	-0.006	0.027	-0.033	-0.002	-0.031	0.033
	(0.436)	(0.016)	(0.016)	(0.022)	(0.021)	(0.017)	(0.023)	(0.024)
Mother's white-collar job	0.106	0.027	0.005	-0.028	0.020	-0.031	-0.017	-0.044
	(0.307)	(0.025)	(0.024)	(0.034)	(0.029)	(0.024)	(0.027)	(0.035)
Household Assets (0-16)	7.59	0.001	0.002	0.003	0.000	0.004	0.003	0.004
	(3.455)	(0.005)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Conventional School	0.245	0.051	-0.066	-0.011	-0.121	0.087	0.086	0.089
	(0.430)	(0.083)	(0.084)	(0.081)	(0.095)	(0.103)	(0.102)	(0.105)
p-value of joint F-test		0.855	0.768	0.720	0.735	0.720	0.325	0.172
Observations		7,957	7,957	3,964	3,993	7,957	3,993	3,964
R-squared		0.070	0.018	0.019	0.030	0.045	0.055	0.041

Notes: This sample consists of 7,971 students who were interviewed at baseline survey. Parent's tertiary education is 1 when they graduate from a 2-year college or a 4-year university. Parent's white-collar job is coded as 1 when they have a professional or government job. Household Assets are defined the total number of assets they own from 16 asset questions. Conventional school is 1 when a student is enrolled in a conventional secondary school. Columns (2) - (8) show randomization balance for three different interventions. Columns (4) and (5) show randomization balance of male circumcision intervention for eligible boys and ineligible girls while columns (7) and (8) show the balance of girls' education support program for eligible girls and ineligible boys, respectively. Robust standard errors clustered by classroom are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

3.3 Student Surveys and Data

The Baseline survey was conducted between January and April 2012 and 7,971 secondary students or 74.4% of total 10,715 students in the school roll-call lists completed the survey. The survey collected information on demographics, household characteristics, education, health, sexual behaviors and preferences in time and risk prior to random assignment. The follow-up survey was conducted approximately 12 months later between January and April 2013 and added cognitive and noncognitive modules on top of the baseline questionnaires. 68.1% of the baseline sample students (or 5,431 students) completed the survey and 31.9% (or 2,540 students) were lost due to attrition because of absence, transfer, or dropout. Since the selectivity of those who attrit could be linked to systematic bias, we randomly chose 15% of the lost students (or 381 students) for tracking (Thomas et al. 2001; Thomas et al. 2012). Out of 381 students, we found out 271 students and conducted the follow-up survey at their home. This resulted in an effective survey follow-up rate of 90.8%.[8] Thus, 5,702 students completed the follow-up survey in total.

Another threat to the experimental design is differential attrition pattern and we investigate whether the likelihood of remaining in the follow-up survey varies by assignment groups and baseline characteristics. Table 3 presents the relationship between survey attrition and baseline characteristics. In the whole sample including ineligible boys, there is no differential attrition bias across assignment groups in columns (1) and (2) as well as across baseline characteristics in columns (3) and (4). However, in the sub-sample of eligible girls only, students who received the scholarship intervention controlling for the full set of demographic characteristics used in Table II were 3.8 percentage points more likely to stay in the sample at the 90% significance level, which causes negative attrition (or retention) bias. We tested whether this attrition is differential by baseline characteristics and found no evidence that the survey attrition due to the girls' education support program is systematically related to baseline characteristics. In order to account for the resulting attrition bias, in all of our regressions we include a whole set of background controls. In addition, in order to examine the robustness of our findings with respect to attrition from the sample, we perform a bounding exercise (Lee 2009) described in detail in footnote 21.

[8] The effective survey rate (ESR) is a function of the regular follow-up rate (RFR) and home-visit follow-up rate (HFR) as follows: ESR = RFR + (1-RFR) * HFR. Overall, ESR is 90.8% (68.1% + 31.9% * 71.1%). Weight for home-visit survey is 6.67 since we did 15% random sampling from the sample attrition (Baird et al. 2012).

The Junior Certificate Examination (JCE) and Malawi School Certificate Examination (MSCE) data were obtained from the District Education Office (DEO) and test scores are normalized so that scores in the control classrooms are distributed with mean 0 and standard deviation 1.[9] JCE and MSCE have 3 core subjects: Chichewa, English, and Math and students have to take at least 3 or 5 additional subjects, respectively. 9^{th} / 11^{th} grade students (at baseline) took JCE / MSCE in June, 2013.

Table 3: Relationship between survey attrition and baseline characteristics

	Whole sample				Eligible girls only			
Dependent variable	= 1 if surveyed in follow-up or home-visit surveys				= 1 if surveyed in follow-up or home-visit surveys			
	Treatment	Adjusted	Main effect	Interaction	Treatment	Adjusted	Main effect	Interaction
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Girls' Education Support	0.006	0.004		0.160	0.040*	0.038*		0.068
	(0.020)	(0.020)		(0.155)	(0.022)	(0.022)		(0.207)
Age		-0.007*	-0.000	-0.014*		-0.015**	-0.012	-0.006
		(0.004)	(0.005)	(0.007)		(0.006)	(0.008)	(0.011)
Orphan		-0.043**	-0.044	-0.003		-0.081**	-0.071	-0.025
		(0.022)	(0.034)	(0.044)		(0.039)	(0.057)	(0.079)
Father's tertiary education		-0.007	-0.025	0.028		-0.003	-0.009	0.001
		(0.017)	(0.026)	(0.034)		(0.025)	(0.038)	(0.051)
Mother's tertiary education		-0.050*	-0.043	-0.012		-0.070**	-0.069	-0.000
		(0.026)	(0.039)	(0.053)		(0.034)	(0.050)	(0.069)
Father's white-collar job		-0.016	-0.019	0.001		-0.024	-0.041	0.035
		(0.012)	(0.018)	(0.025)		(0.017)	(0.027)	(0.034)
Mother's white-collar job		-0.012	0.016	-0.054		-0.028	-0.004	-0.049
		(0.022)	(0.031)	(0.041)		(0.027)	(0.038)	(0.053)
Household Assets		-0.001	-0.004	0.006		-0.001	-0.005	0.006
		(0.002)	(0.003)	(0.005)		(0.003)	(0.004)	(0.007)
Conventional School		0.053***	-0.002	0.100**		0.049**	-0.008	0.102**
		(0.019)	(0.024)	(0.039)		(0.023)	(0.035)	(0.048)
Observations	7,971	7,957	7,957		3,997	3,993	3,993	
R-squared	0.011	0.016	0.020		0.014	0.024	0.027	

Notes: Columns (3) and (4) present results from one regression with main effects (column 3) and all covariates interacted with treatment effect (column 4). Regressions are OLS models with grade fixed effects. Robust standard errors clustered by classroom are reported in parenthesis. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

[9] The JCE / MSCE data contain three identifying variables for each student: attending school, gender, and name. However, an exact match on these identifying variables is very difficult because of differences in the spelling of names. To deal with the issue of differently spelled names we applied probabilistic matching algorithms in the spelling of names by using reclink STATA command (de Hoop, 2011). The matching algorithm provides a match score indicating how closely two names match on a scale from 0 to 1. We used a cutoff with a match score below 0.6 (by default) and we have been conservative in the matching procedure by checking all approximate string matches manually.

4 Estimation Strategy

The random assignment of girls education support program allows us to ensure that the assignment of the treatment is orthogonal to other characteristics of the sample that may be correlated with outcome variables. We focus on reduced-form estimation of the program impacts on schooling outcomes and exam scores. To better understand possible mechanisms underlying the impacts on schooling and examinations, we also evaluate program impacts on several channels, including child labor, health, cognitive/noncognitive abilities and time preferences by estimating the following simple model using ordinary least squares:[10]

$$Y_{ic} = \beta_0 + \beta_1 ES_c + \beta_2 X_{ic} + \delta_c + \varepsilon_{ic} \tag{1}$$

The variable $Y_{ic}$is the outcome for student i, in classroom c. The variable ES_c is an indicator for whether classroom c was assigned for education support program and the coefficient β_1 captures the average program impact. X_{ic} is a vector that includes the sociodemographic controls.[11] We also include grade fixed effects, δ_c since the randomization was implemented within grade. In some specifications we also control for the effect of the other two interventions (HIV/AIDS education and circumcision for boys) and their interactions, since they were part of the broader project.

5 Results

5.1 Schooling: Dropout

Table 4 describes self-reported and school-reported dropout rates.[12] Both are in similar magnitude but insignificantly different from zero. When we examine dropout and transfer together in columns

[10]When outcome variables are binary, we have also estimated the main treatment effects using probit models, yielding consistent results.

[11]We included age, orphan status, parents tertiary education, parents white-collar job, household asset ownership, and school type in order to address any minor baseline differences between the treatment and control classrooms that exist despite the randomization. Since the randomization successfully produced treatment and control classrooms balanced across most baseline characteristics, the inclusion of these controls does not significantly change the treatment effect estimates but does sometimes improve statistical precision.

[12]Kremer, Miguel, and Thornton (2009) use school participation data based on unannounced checks by NGO enumerators whereas Baird, McIntosh, and Özler (2011) use school enrollment and attendance information from official school ledgers as benchmark while collecting self-reported and teacher-reported data as well. We collected both self-reported and official school attendance data. School-reported dropout and transfer were consistent with self-reported data in the sample. 82% and 84% of school-reported dropout and transfer were exactly matched with self-reported data.

(5) and (6), impact estimates suggest that the probability of being enrolled in the original school increase by 8.3 percentage points at the 99% confidence level. Since not all students who transferred to other schools were actually enrolled, this estimate can be regarded as the upper bound for school dropout. Panel B, C, and D show differential dropout pattern by grade and it is suggested that dropout rates decrease as students advance higher grades. We observe significant decrease in school dropout among 9^{th} grade, which is starting grade of a secondary school, while the girls education support program does not affect 10^{th} and 11^{th} grade students school dropout rates.[13] To examine the robustness of our results, we report the coefficient both with and without the baseline controls. The inclusion of these controls does not significantly affect the estimates, as expected given that the covariates are balanced across classrooms. This robustness check approach is applied in all specifications.

[13]When secondary-school-aged girls are not in school, they are likely to be unemployed, to get married, and to begin childbearing according to Malawi Demographic and Health Survey 2010. 63.4% of females whose ages are 15-19 years old are unemployed and only 36.5% are employed (most of female employments are agriculture farming and domestic work which do not have earned wages). 23.4% of those girls in the same age bracket get married or start cohabiting while 26% of them begin childbearing. These figures on marriage and childbearing rapidly increase when these teenage girls enter the 20-24 age bracket.

Table 4: Impact on School Dropout

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	drop-out (self-reported)		drop-out (school-reported)		drop-out or transfer (school-reported)	
Panel A: Whole sample						
Girls' Education Support	-0.039	-0.034	-0.032	-0.027	-0.087***	-0.083***
	(0.029)	(0.023)	(0.024)	(0.021)	(0.033)	(0.030)
Observations	2,861	2,860	2,861	2,860	2,861	2,860
Panel B: Grade 9						
Girls' Education Support	-0.101**	-0.090**	-0.060*	-0.044*	-0.124**	-0.127***
	(0.049)	(0.037)	(0.033)	(0.024)	(0.046)	(0.031)
Observations	889	889	889	889	889	889
Panel C: Grade 10						
Girl's Education Support	0.012	0.018	0.023	0.018	-0.051	-0.084**
	(0.047)	(0.030)	(0.050)	(0.036)	(0.055)	(0.041)
Observations	1,040	1,039	1,040	1,039	1,040	1,039
Panel D: Grade 11						
Girls' Education Support	-0.045	-0.017	-0.074**	-0.059*	-0.099	-0.087
	(0.050)	(0.049)	(0.033)	(0.035)	(0.068)	(0.066)
Observations	932	932	932	932	932	932
Controls	No	Yes	No	Yes	No	Yes

Notes: The dependent variable in Columns (1) and (2) are based on follow-up survey while that in Columns (3) and (4) are based on difference between baseline and follow-up surveys. Regressions are OLS models with grade fixed effects. Robust standard errors clustered by classroom are reported in parenthesis. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls: age, orphan status, parents' tertiary education, parents' white-collar job, household assets, and school type. *** p<0.01, ** p<0.05, * p<0.10

5.2 Schooling: Attendance

Although we didnt find the extensive margin on school dropout (or enrollment) overall, Table 5 indicates that the program has an effect on the intensive margin by reducing school absence for those enrolled in school. The average absences based on self-reported data was 3.8 days per semester (or 11.4 days per year) and treated girls are 1.6 days per semester (or 5 days per year) less likely to be absent.[14] When we control baseline school absence level, we have slightly bigger impacts shown in columns (3) and (4).

[14]Baird, McIntosh, and Özler (2011) reported that conditional cash transfers increased school attendance by 10 more school days over the entire school year. The magnitude of our results on school absence (5 more school days) is half of that of Baird, McIntosh, and Özler (2011) but our girls education support program didnt impose school attendance conditionality.

Table 5: Impact on School Attendance

Dependent variable	Self-reported absence			
	(1)	(2)	(3)	(4)
Girls' Education Support	-1.707***	-1.645***	-2.187***	-2.187***
	(0.345)	(0.275)	(0.437)	(0.420)
Mean in the control group	3.794		0.672	
Controls	No	Yes	No	Yes
Observations	2,715	2,704	2,700	2,689
R-squared	0.027	0.047	0.027	0.042

Notes: The dependent variable in Columns (1) and (2) are based on follow-up survey while that in Columns (3) and (4) are based on difference between baseline and follow-up surveys. Regressions are OLS models with grade fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

5.3 Test Scores: Malawi School Certificate Examination (MSCE)

In Table 6, we present the impact of scholarship program on the performance on the Malawi School Certificate Examination (MSCE) for 11^{th} grade students at baseline. A first outcome we consider is simply whether students took the MSCE.[15] Column (2) suggests that students in the treatment classrooms were 2 percentage points less likely to take the MSCE but this result is statistically insignificant. Though the point estimate is negative, the upper bound of the 95% confidence interval is 7.8 percentage points increase in taking MSCE. Since there was no evidence of selection into exam taking, we were able to simply interpret effects on MSCE performance. Only 34.7% of the sample passed the MSCE and the remaining 65.3% of the students graduated without MSCE certificate.[16] Columns (3) and (4) report that there was no significant improvement on the probability to pass MSCE though the point estimates are positive and columns (5) - (12) show no impact on overall scores and the three core subjects: Chichewa, English, and Mathematics.[17] One reason of no impact on MSCE may be the relatively short period between baseline and follow-up surveys. To the extent that improving exam scores especially in a difficult testing takes time, our

[15] When MSCE and JCE data from Division Education Office are matched with our baseline sample, we define this match as exam taking. There are 1,186 11^{th} grade female students in the sample and MSCE data are matched with 74.1% (or 879 students) of the sample.

[16] We code MSCE passing variable as 0 for students whose MSCE data from Division Education Office are missing or are not matched with the baseline survey. If we define MSCE passing variable for only those who took MSCE (or whose MSCE data is matched), then 46.8% of the sample passed the exam.

[17] The overall score is the standardized sum of three core subject scores (Chichewa, English, and Math), each standardized on its own before the sum. Thus, the overall score has mean 0 and standard deviation of 1 of the control group.

current 1-year short term follow-up may not be able to fully capture the full range of possible effects. This hypothesis can be tested when we obtain MSCE data for baseline 9^{th} and 10^{th} grade students in the following years.

Table 6: Impact on Malawi School Certificate Exam (MSCE) 2013 for 11^{th} grade students

VARIABLES	= 1 if took MSCE		= 1 if passed MSCE		Overall score		Chichewa score		English score		Math score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Girls' Education Support	-0.026	-0.020	0.038	0.023	0.054	-0.013	0.074	0.007	0.085	0.030	-0.020	-0.076
	(0.055)	(0.048)	(0.069)	(0.061)	(0.169)	(0.134)	(0.195)	(0.158)	(0.181)	(0.125)	(0.139)	(0.131)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,186	1,186	1,186	1,186	851	851	853	853	867	867	865	865
R-squared	0.009	0.028	0.002	0.088	0.001	0.122	0.002	0.101	0.002	0.193	0.000	0.052

Notes: MSCE scores: 0 - fail, 1 2 - pass, 3 6 - credit, 7 8 - distinction. Chichewa, English, and Math are three core subjects. Overall and three core subject scores have been standardized to have a mean of 0 and a standard deviation of 1 in the control group. Regressions are OLS models with school district fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** $p<0.01$, ** $p<0.05$, * $p<0.10$

5.4 Test Scores: Junior Certificate Examination (JCE)

Table 7 shows the program impact on the Junior Certificate Examination (JCE) for 9^{th} grade students at baseline. Treated girls were 15.5 percentage points more likely to take the JCE at the 99% significance level in column (2) and this result is consistent with the finding that 9^{th} grade students were more likely to stay in school, which translates into the increased probability of taking the exam one year later. We also see the improvements in JCE performances.[18] The probability of passing the JCE improved by 18.7 percentage points at the 99% confidence level and the overall scores increased by 0.241 SD at 90% level when we adjusted the treatment effect with the baseline characteristics. The improvement on the overall scores attributed to the increased scores in Chichewa subject but no significant improvement can be detected in English and Math subjects. However, these results should be taken cautiously due to the selection bias in the first stage of JCE taking. Because a higher proportion of students from treatment classrooms took the JCE, it is likely that the group of students from treatment classrooms contains a higher fraction of relatively weak students; that is, strong students are likely to take the JCE regardless of the scholarship program, but marginal students who are induced to take the exam because they received scholarship are likely to be relatively lower scoring students. Such a selection process could bias downward the effect of scholarship program on the average JCE performances we report in Table 1.7.[19]

[18]Compared to MSCE, JCE is relatively easier to pass. 59.8% of the sample passed the exam. If we restrict the sample to those who took the exam, then 84.4% of students who took JCE passed the exam. The reason why we have impact on JCE but not on MSCE could be thought in that JCE is a low-stakes testing and MSCE is a high-stakes testing. However, both JCE and MSCE are high-stakes tests because they have important consequences for students grade promotion and graduation with certificate. We were unable to analyze low-stakes tests such as regular assessment exams from each school due to huge missing data and incompatibility across the target schools.

[19]Krueger and Whitmore (2001) reported the impact of the Tennessee STAR experiment on the ACT or SAT college entrance exam. They show that attending a small class in the early grades is associated with an increased likelihood of taking a college-entrance exam but no significant improvements in test scores can be observed arguably due to the selection bias on the extensive margin of exam taking. They use Hackman selection correction and linear truncation model to adjust the selection issue under the assumption that the same factors that determine whether student are more likely to take the exam also determine how well they do on the test, After correction the selection problem through either Hackman selection or linear truncation models, they found significant improvements in the exam scores.

Table 7: Impact on Junior Certificate Exam (JCE) 2013 for 9^{th} grade students

VARIABLES	= 1 if took JCE		= 1 if passed JCE		Overall score		Chichewa score		English score		Math score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Girls' Education Support	0.173***	0.155***	0.151**	0.187***	0.196	0.241*	0.226	0.261**	0.094	0.193	0.156	0.143
	(0.037)	(0.038)	(0.060)	(0.048)	(0.188)	(0.127)	(0.153)	(0.128)	(0.214)	(0.137)	(0.151)	(0.108)
Controls	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,220	1,220	1,220	1,220	853	853	856	856	855	855	858	858
R-squared	0.041	0.052	0.023	0.091	0.010	0.242	0.014	0.120	0.002	0.324	0.005	0.142

Notes: JCE scores: 0 - fail, 1 - average, 2 - good, 3 - very good, 4 - excellent. Chichewa, English, and Math are three core subjects. JCE scores are standardized with mean 0 and standard deviation of 1 of the control group. Regressions are OLS models with school district fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

5.5 Cognitive Ability

We now turn to an analysis of the cognitive ability. The cognitive test score is based on 30 Ravens Matrices (18 questions from Colored Progressive Matrices and 12 questions from Standard Progressive Matrices). In Table 1.8, we see improvements of 0.215 SD in the treatment classrooms compared with the control classrooms at the 99% confidence level. When the demographic controls are included, the coefficient on cognitive ability remains similar (0.2 SD). This effect size of more than 0.2 SD is not only statistically but also economically significant compared to similar program effect size such as Baird, McIntosh, and Özler (2011) reporting that the impact of conditional cash transfers for secondary female students in Malawi on cognitive test score (based on Ravens Matrices) was 0.17 SD increase.[20]

Since we have survey attrition bias of 3.9 percentage points at the 90% level presented in Table 3 despite the fact that there is no differential attrition across baseline characteristics, we perform bounding exercises by assigning higher percentile scores to attrition students belong to the control classroom while assigning lower percentile scores to attrition students belong to the treatment classroom.[21] As an example, among the representative attrition sample of 187 students (15% random sampling of total 2,540 attrition students), 59 girls missed the follow-up survey (28 in the treatment group and 31 in the control group). The 90-percentile score was assigned to 31 girls in the control group while 28 girls in the treatment group were assigned the 10-percentile score. Columns (3) and (4) show that even after a bounding exercise, the findings on treatment impacts

[20]Cohen (1969) describes an effect size of 0.2 SD as the difference between the heights of 15 year old and 16 year old girls in the US and an effect size of 0.5 SD corresponds to the difference between the heights of 14 year old and 18 year old girls. In education, if an intervention would raise academic outcomes by an effect size of even as little as 0.1 SD, then this could be a very significant improvement, particularly if the improvement applied uniformly to all students, and even more so if the effect were cumulative over time. Other education interventions such as building a village-based primary school in Afghanistan (Burde and Linden 2013) and girls merit-based scholarship in Kenya (Kremer, Miguel, and Thornton 2009) reported 0.65 SD and 0.19 SD increases on average test scores. An effect size of 0.65 SD from the Afghanistan primary school project is considered the largest one among many education field interventions conducted by J-PAL (Jameel Poverty Action Lab) at MIT.

[21]Among the representative attrition sample of 187 students, 59 girls have missing follow-up survey (28 in the treatment group and 31 in the control group). Initially, 59 attrition girls were assigned 50 percentile score and then we increased the percentile score assigned to 31 girls in the control group by 1 percentile while decreasing the percentile score to 28 girls in the treatment group by 1 percentile. Therefore, the second bounding practice is that 31 girls in the control group were assigned 51 percentile score and 28 in the treatment 49 percentile score. In the end, we assigned 99 percentile score to the control group girls and 1 percentile score to the treatment group girls. The findings on treatment impacts would not change until we assigned 93 percentile and 7 percentile scores to the attrition when we run the basic regression without control variables. If we used the specification with control variables, then the treatment impact would hold even after 99 percentile and 1 percentile scores were assigned to the attrition students (Appendix 2).

would not change, even though the size of the coefficient decreases from 0.215 SD to 0.166 SD.

Table 8: Impact on Cognitive Ability

Dep. Var.	Cognitive Test Score		Cognitive Test Score (bounding practice)	
	(1)	(2)	(3)	(4)
Girls' Education Support	0.215***	0.200***	0.166**	0.155**
	(0.082)	(0.065)	(0.082)	(0.064)
Controls	No	Yes	No	Yes
Observations	2,861	2,850	2,920	2,908
R-squared	0.021	0.132	0.016	0.126

Notes: The cognitive test score is based on 30 Raven's Matrices (18 questions from Colored Progressive Matrices and 12 questions from Standard Progressive Matrices). The test scores have been standardized to have a mean of 0 and a standard deviation of 1 in the control group. Columns (3) and (4) show a bounding practice. Among the representative attrition sample of 187 students, 59 girls have missing follow-up survey (28 in the treatment group and 31 in the control group). 90 percentile score was assigned to 31 girls in the control group while 28 girls in the treatment group were assigned 10 percentile score. Regressions are OLS models with grade fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls in the regression analyses: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

We also take into consideration the fact that the test settings for cognitive ability were different between follow-up and home-visit surveys. Students who participated in the follow-up survey took the cognitive test at their school while those who missed the follow-up survey and later completed the home-visit survey took the test mostly at their home.[22] In order to adjust the setting difference, we randomly selected 2% of the follow-up sample and visited them again and administered the cognitive ability test with additional 5 new questions at their home. Through difference-in-difference estimation, we find no evidence of the test setting effect on whether it is administered at school or at home (Appendix 3).

[22] Duflo, Dupas, and Kremer (2011) brought attrition students to school for the test while Baird, McIntosh, and Özler (2011) administered the test to all study participants at their homes.

5.6 Noncognitive Ability

We measured four variables on noncognitive abilities: self-esteem, self-efficacy, aspiration for education, and importance of education. Self-esteem and self-efficacy were measured by 6 and 4 related questionnaires, respectively while aspiration for education is a dummy variable when students aim to continue on to bachelor or master degrees. Importance of education is a categorical data ranging from 1: not important at all to 5: very important. Table 9 shows the program impact on noncognitive abilities. Although we do not observe significant change in self-esteem and self-efficacy, we find that the scholarship program resulted in improvements in aspirations for education and importance of education.[23] When controlled baseline characteristics, educational aspirations of the treatment classrooms increased by 0.127 SD in column (6) at the 90% significant level. Importance of education was also improved by 0.112 SD in column (8) at the 95% significant level. The improvements in aspirations and importance of education were partly supported by the result that students in the treatment classrooms were more likely to study after school (not shown).

[23]One concern from psychology argues that extrinsic rewards like scholarship or cash transfers may interfere with intrinsic motivation and could thus reduce effort in some circumstances. Surveys of students in our Malawian data provide no evidence that scholarship or cash transfers weaken intrinsic motivation for education at least in the short-run.

Table 9: Impact on Non-cognitive Abilities

Dependent Variable	Self-esteem		Self-efficacy		Aspirations for education		Importance of education	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Girls' Education Support	0.003	0.005	-0.045	-0.034	0.152	0.144**	0.099**	0.112**
	(0.075)	(0.063)	(0.069)	(0.055)	(0.093)	(0.072)	(0.045)	(0.048)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,861	2,860	2,861	2,860	2,861	2,860	2,799	2,798
R-squared	0.013	0.054	0.017	0.057	0.006	0.102	0.003	0.007

Notes: Noncognitive abilities have been standardized to have a mean of 0 and a standard deviation of 1 in the control group. Self-esteem was constructed by 6 questions while self-efficacy is based on 4 related questions. Aspiration for education is coded as 1 when students aim to study 4-year university or master's degree. Importance of education (1: Not important at all 5: Very Important) is based on the difference between baseline and follow-up survey responses. All coefficients are standardized to have a mean of 0 and a standard deviation of 1 in the control group. Regressions are OLS models with grade fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls in the regression analyses: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** $p<0.01$, ** $p<0.05$, * $p<0.10$

5.7 Time Preference

Recent research has shown that experimentally elicited time preference of adolescents is a significant predictor on alcohol, cigarettes, BMI, saving, and proper school conduct (Sutter et al. 2013). Since time preference seems more directly related to self-control, which is an important mechanism for educational achievement, we elicit time preference in our baseline and follow-up surveys using simple versions of hypothetical choice lists that are widely used in the economics literature.[24] Table 10 shows hypothetical choice lists used in the baseline and follow-up surveys. Panel A presents the original format of choice lists for time preference. In the first question (Q801), students choose either 10,000 MK (US $57) in 1 month or 20,000 MK (US $114) in 3 months.[25] If 10,000 MK in 1 month is chosen, then students move to Q802 and select three choice lists with increasing payoffs whereas the other students who choose 20,000 MK in 3 months go to Q803 and select three choice lists with decreasing payoffs. In panel B, we rearranged the choice lists with monotonically increasing payoffs as if students were asked to start choosing choices from the first list to the ninth one. From the rearranged lists we calculated the future equivalent of the fixed payoff at the earlier point in time as the midpoint between the two later payoffs where a student switches from the earlier to the later payment.[26] For example, if a student ends up with the second choice list (11,000 MK in 3 months over 10,000 MK in 1 month) which is the most time-patient category in our setting, then the future equivalent equals 10,500MK. If students choose the ninth choice list (10,000 MK in 1 month over 50,000 MK in 3 months) which is the most time-impatient category in our setting, then the future equivalent cannot be calculated and is censored by Tobit regressions.

[24]The use of real rewards is generally desirable, but hypothetical rewards actually have some advantages. In studies involving hypothetical rewards, respondents can be presented with a wide range of reward amounts, including losses and large gains, both of which are generally infeasible in studies involving real outcomes. The disadvantage of hypothetical choice data is the uncertainty about whether people are accurately predicting what they would do if outcomes were real.

[25]Instead of using choices between today and three months, we used choices between 1 month and 3 months in order to rule out present-biased discounting. Sutter et al. (2013) reported no strong evidence for present-biased discounting for 661 adolescents, aged 10 to 18 years in Austria.

[26]Just like Sutter el al. (2013), we calculate the discount rate with i = ln (future equivalent / initial payoff (=10,000 MK)) x 12/2 in case of a two-month delay.

Table 10: Choice Lists for Time Preference

Panel A: Original Choice Lists				
Q801		0 = 10,000 MK in 1 month (Go to Q802)	or	1 = 20,000 MK in 3 months (Go to Q803)
	[1]	0 = 10,000 MK in 1 month	or	1 = 30,000 MK in 3 months
Q802	[2]	0 = 10,000 MK in 1 month	or	1 = 40,000 MK in 3 months
	[3]	0 = 10,000 MK in 1 month	or	1 = 50,000 MK in 3 months
	[1]	0 = 10,000 MK in 1 month	or	1 = 17,000 MK in 3 months
Q803	[2]	0 = 10,000 MK in 1 month	or	1 = 14,000 MK in 3 months
	[3]	0 = 10,000 MK in 1 month	or	1 = 11,000 MK in 3 months
Panel B: Rearranged Choice Lists with monotonically increasing payoffs				
	[1]	0 = 10,000 MK in 1 month	or	1 = 10,000 MK in 3 months
	[2]	0 = 10,000 MK in 1 month	or	1 = 11,000 MK in 3 months
	[3]	0 = 10,000 MK in 1 month	or	1 = 14,000 MK in 3 months
	[4]	0 = 10,000 MK in 1 month	or	1 = 17,000 MK in 3 months
	[5]	0 = 10,000 MK in 1 month	or	1 = 20,000 MK in 3 months
	[6]	0 = 10,000 MK in 1 month	or	1 = 30,000 MK in 3 months
	[7]	0 = 10,000 MK in 1 month	or	1 = 40,000 MK in 3 months
	[8]	0 = 10,000 MK in 1 month	or	1 = 50,000 MK in 3 months
	[9]	0 = 10,000 MK in 1 month	or	undefined

Notes: Panel A shows the initial questionnaire structure for time preference module. First, students are asked whether they prefer 10,000 Malawi Kwacha in 1 month to 20,000 Malawi Kwacha in 3 months. If students choose 10,000 MK in month, then they move to three sequence questions in Q802. If they choose 20,000 MK in 3 months, then they go to three sequence questions in Q803 instead of Q802. Panel B rearranges choice lists in Panel A as payoffs are monotonically increasing. The fifth choice list in Panel B is the starting question of Q801 for time preference module.

Table 11 shows that time preference at the baseline is significantly correlated with gender (females are more patient) whereas there is no age effect in time preference, with such findings being consistent with Sutter et al. (2013) and Castillo et al. (2011). However, the result that more household assets are correlated to higher discount rate is different from Tanaka, Camerer, and Nguyen (2010) showing how higher household income is positively correlated to time patience in Vietnamese villages.

Table 11: Annual Discount Rates and Baseline Characteristics

Dep. Var.	Discount rate
Female	-0.268***
	(0.060)
Age	0.003
	(0.017)
Orphan	-0.011
	(0.125)
Father's tertiary education	-0.121
	(0.087)
Mother's tertiary education	0.019
	(0.130)
Father's white-collar job	-0.039
	(0.074)
Mother's white-collar job	-0.107
	(0.114)
Household asset ownership	0.021**
	(0.009)
Observations	7,905

Notes: Table 11 shows the correlations between baseline annual discount rates and baseline characteristics. Regression is OLS model. Robust standard errors clustered by classroom are reported in parentheses. Baseline values of the following variables are included as controls in the regression analyses: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

Table 12 presents interesting evidence that one-year education support program improved time patience among girls at secondary schools. The program decreased the annual discount rate by 36.7 percentage points on average and the average impact falls to 33.9 percentage points when the baseline controls are included in the regression. Since the average of annual discount rate at the follow-up survey was 538%, the program impact would be a 0.123 SD - 0.137 SD decrease in the annual discount rate.[27] This may seem like a small effect but improved time preference can affect many dimensions on education achievement, health behaviors, and employment and may accumulate to substantial effects in total in the long run.

[27] The average annual discount rate of 538% for the delay of two months may seem to be considerably large. However, it is common to assume that the future discount rates in developing countries are quite high. Kohler and Thornton (2012) evaluate the impact of financial incentives (conditional cash transfers) for individuals in Malawi to maintain their HIV status for one year but find no statistical difference in reported behaviors. They consider that the offer of the financial reward one year later was too far away from the present to overcome hyperbolic discounting. Moreover, compared to Sutter et al. (2013) reporting that the annual discount rates for the delay of three weeks among 661 adolescents aged 10-18 years in Austria are 330% and 365% (with the three-week upfront-delay), annual discount rate of 538% among teenage girls in Malawi can be considered not unusual.

Table 12: Impact on Time Preference

Dependent Variable	Discount rate	
	(1)	(2)
Girls' Education Support	-0.367**	-0.339**
	(0.144)	(0.140)
Mean of Dep. Var. in control group	5.38	
Controls	No	Yes
Observations	2,822	2,821

Notes: The dependent variable is annual discount rates. Regressions are Tobit model with grade fixed effects. Robust standard errors clustered by classroom are reported in parentheses. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls in the regression analyses: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10

5.8 Heterogeneity of Program Impacts and Spillovers to boys

We now turn to an examination of impact heterogeneity by a number of baseline characteristics: orphan status, parents education, parents job, and household asset index. We do not find significant impact heterogeneities across outcome variables from schooling outcomes to time preferences (not shown). We also tested whether there were spillover effects to ineligible boys in the treatment classrooms. Unlike Kremer, Miguel, and Thornton (2009) who find a spillover effect of a merit-based girls scholarship program to ineligible boys in the same classroom on test scores, we do not find significant spillovers to boys on most of outcome variables (not shown).

6 Conclusion

The impacts of education have traditionally been measured by test scores and underlying mechanism for test score gains has been explained by simple schooling outcomes such as enrollment and attendance. This paper investigates the impacts of girls' education support program among secondary students not only on schooling outcomes and test scores but also on cognitive/non-cognitive abilities and time preferences using data from a randomized controlled trial in Malawi.[28] Students who were randomly assigned to the program are less likely to be dropped out and to be absent from school. Although there are improvements in the Junior Certificate Examination scores for

[28]Since we havent included related male classmates, teachers, and parents of the target female students in the baseline survey, we are unable to analyze the general equilibrium effects of the girls education support program in a larger context. Thus, all the results reported in this paper are only partial equilibrium effects of the intervention.

grade 9 students, no significant impacts in the Malawi Certificate Examination for grade 11 students can be detected. However, we find evidence that the treated students do much better on cognitive test and show modest improvements on noncognitive abilities. They also become more time patient. Researcher who had examined only test scores would have incorrectly concluded that the education support program does not have a significant impact. Since cognitive/non-cognitive abilities and preferences are considered to be more stable and persistent than test scores, our results on improvements of those outcomes may suggest that future educational achievement as well as adulthood labor market outcomes can be affected by these improved cognitive/non-cognitive abilities and preferences. Further follow-up research using this sample of students in the coming years will try to shed light on how persistent the impacts of girls' education support program on cognitive/non-cognitive abilities and preferences are and thus will help us understand how these mechanism shape socio-economic outcomes during adulthood.

References

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92 (5):1535–1558.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96 (3):847–862.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2012. "Worms at Work: Public Finance Implications of a Child Health Investment." *Unpublished Working Paper* .

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *Quarterly Journal of Economics* 126 (4):1709–1753.

Barnett, W Steven. 1992. "Benefits of Compensatory Preschool Education." *Journal of Human Resources* 27 (2):279–312.

Behrman, Jere R, Piyali Sengupta, and Petra Todd. 2005. "Progressing through PROGRESA: an Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change* 54 (1):237–275.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2013. "Turning a Shove into a Nudge? A Labeled Cash Transfer for Education." *NBER Working Paper* (w19227).

Blau, Francine D and Lawrence M Kahn. 2005. "Do Cognitive Test Scores Explain Higher US Wage Inequality?" *Review of Economics and Statistics* 87 (1):184–193.

Burde, Dana and Leigh L Linden. 2013. "Bringing Education to Afghan Girls: A Randomized Controlled Trial of Village-based Schools." *American Economic Journal: Applied Economics* 5 (3):27–40.

Castillo, Marco, Paul J Ferraro, Jeffrey L Jordan, and Ragan Petrie. 2011. "The Today and

Tomorrow of Kids: Time Preferences and Educational Outcomes of Children." *Journal of Public Economics* 95 (11):1377–1385.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR." *Quarterly Journal of Economics* 126 (4):1593–1660.

Cohen, Jacob. 2013. *Statistical power analysis for the behavioral sciences*. Routledge Academic.

De Hoop, Jacobus. 2011. "Selective Schools and Education Decisions: Evidence from Malawi." *Unpublished working paper, Free University of Amsterdam* .

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101:1739–1774.

Duflo, Esther, Pascaline Dupas, Michael Kremer, and Samuel Sinei. 2006. "Education and HIV/AIDS Prevention: Evidence from a Randomized Evaluation in Western Kenya." *World Bank Policy Research Working Paper 4024* .

Glewwe, Paul, Phillip H. Ross, and Bruce Wydick. 2013. "Developing Aspirations: The Impact of Child Sponsorship on Self-Esteem and Life Expectations." *Unpublished working paper* .

Hanushek, Eric A and Ludger Woessmann. 2012. "Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation." *Journal of Economic Growth* 17 (4):267–321.

Kremer, Michael, Edward Miguel, and Rebecca Thornton. 2009. "Incentives to Learn." *Review of Economics and Statistics* 91 (3):437–456.

Krueger, Alan B and Diane M Whitmore. 2001. "The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR." *Economic Journal* 111 (468):1–28.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76 (3):1071–1102.

Macours, Karen, Norbert Schady, and Renos Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* 4 (2):247–273.

National Statistical Office (NSO). 2011. *Welfare Monitoring Survey.* Lilongwe, Malawi: Government of Malawi (GOM).

Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexiacan Progresa Poverty Program." *Journal of Development Economics* (74):199–250.

Son, Hyelim. 2013. "Human Capital Investment When Sheepskin Effects Matter: Evidence from Income Shocks in Indonesia." *Job Market Paper, Department of Economics, Columbia University* .

Sutter, Matthias, Martin G Kocher, Daniela Glatzle-Rutzler, and Stefan T Trautmann. 2013. "Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior." *American Economic Review* 103 (1):510–531.

Tanaka, Tomomi, Colin F Camerer, and Quang Nguyen. 2010. "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam." *American Economic Review* 100 (1):557–571.

Thomas, Duncan, Elizabeth Frankenberg, and James P Smith. 2001. "Lost but Not Forgotten." *Journal of Human Resources* 36 (3):556–592.

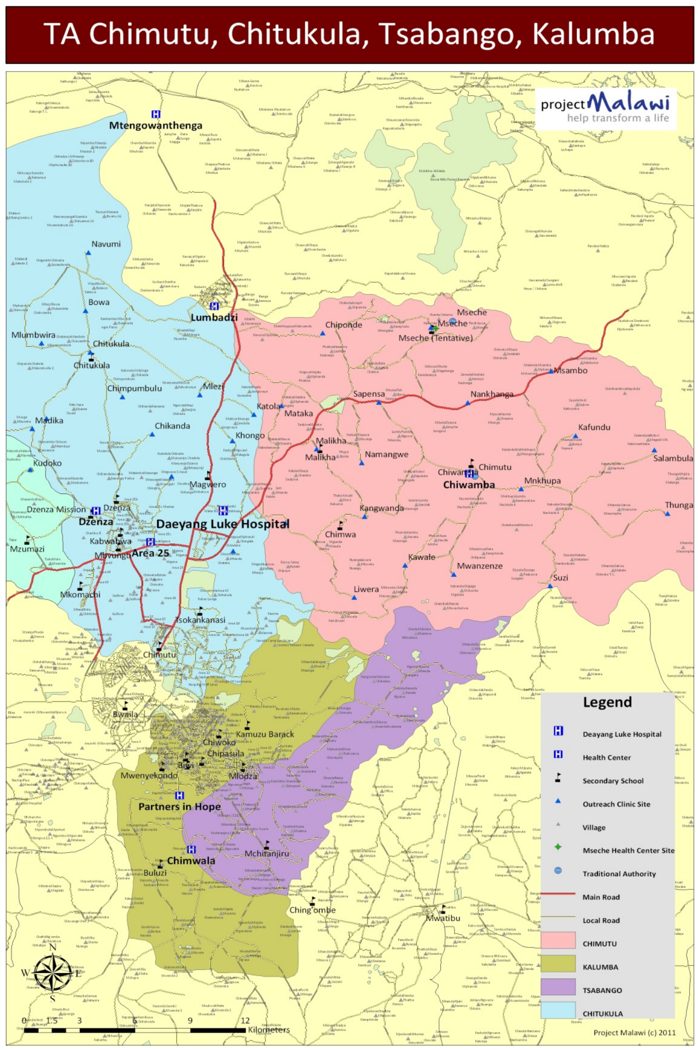
Thomas, Duncan, Firman Witoelar, Elizabeth Frankenberg, Bondan Sikoki, John Strauss, Cecep Sumantri, and Wayan Suriastini. 2012. "Cutting the costs of attrition: Results from the Indonesia Family Life Survey." *Journal of Development Economics* 98 (1):108–123.

Wydick, Bruce, Paul Glewwe, and Laine Rutledge. 2013. "Does international child sponsorship work? A six-country study of impacts on adult life outcomes." *Journal of Political Economy* 121 (2):393–436.

Appendix

Appendix 1: Map of Project Areas

Figure 1: Map of Project Areas

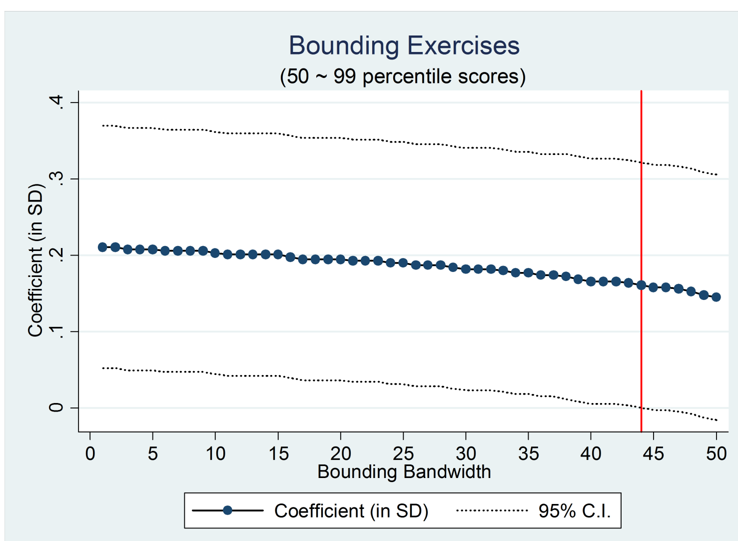


Appendix 2: Bounding Exercises for Cognitive Ability

Initially, we assigned 50 percentile cognitive test score to the attrition sample and then we increased the percentile score assigned to attrition students of the control group by 1 percentile while decreasing the percentile score to attrition students of the treatment group by 1 percentile. Therefore, the second bounding practice is that attrition students in the control group was assigned 51 percentile score and those in the treatment group 49 percentile score. In the end, we assigned 99 percentile score to the control group attrition and 1 percentile score to the treatment group attrition.

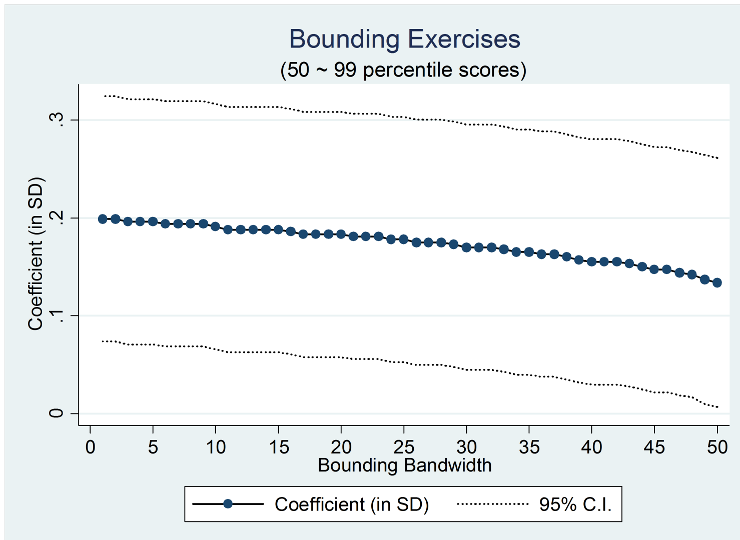
1) Regression without baseline controls: The findings on cognitive ability would not change until we assigned 93 percentile and 7 percentile scores to the attrition.

Figure 2: Bounding excercises 1



2) Regression with baseline controls: The findings on cognitive ability would not change even when we assigned the highest 99 percentile and the lowest 1 percentile scores to the attrition.

Figure 3: Bounding excercises 2



This figure plots the change of coefficients on cognitive ability in Table 8 with different bounding exercises. To construct these figures, we bin bounding exercises into fifty different bounding bandwidths and plot the standard deviation of cognitive ability for each bandwidth. The dashed lines show upper and lower intervals of the 95% confidence level.

Appendix 3: Test Setting Effect

The settings for cognitive ability test were different between follow-up and home-visit surveys. Table 13 presents that follow-up survey sample took the cognitive test at their school while the test was administered mostly at home for home-visit survey. The impact on cognitive ability in Table 8 was derived from both 30 questions of follow-up survey sample (1) and home-visit survey sample (3), which does not adjust the setting difference.

Table 13: Setting Effect of Cognitive Score

		Follow-up survey sample	Home-visit survey sample
Test Location	School	(1) 30 questions	N/A
	Home	(2) 30 questions (4) 5 questions	(3) 30 questions (5) 5 questions

In order to adjust the test setting difference, we randomly select 2% of the students who completed follow-up survey and visited them again and administered the cognitive ability test with additional 5 new questions at their home. Table 14 shows that there is no evidence of significant test setting effect on cognitive ability through diff-in-diff estimation.

Table 14: Setting adjustment: diff-in-diff estimation

	Cognitive Test Score	
	(1)	(2)
30 questions vs. 5 questions	-0.022	-0.022
(1) - (4) or (3) - (5)	(0.021)	(0.022)
follow-up vs. home-visit	0.071	0.073
(1) - (3) or (4) - (5)	(0.045)	(0.048)
30 questions x follow-up	-0.005	-0.005
((1)-(4)) - ((3)-(5))	(0.043)	(0.044)
Controls	No	Yes
Observations	370	370
R-squared	0.044	0.192

Notes: 2% random sample for follow-up survey and home-visit survey samples were weighted with 50 and 6.67, respectively. Regressions are OLS models with grade fixed effects. Robust standard errors are clustered by classroom. The weight of 6.67 is given to home-visit survey sample. Baseline values of the following variables are included as controls in the regression analyses: age, orphan status, parents' tertiary education, parents' white-collar job, household asset ownership, and school type. *** p<0.01, ** p<0.05, * p<0.10