INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

IFPRI

IFPRI Discussion Paper 01627

March 2017

Participation, Learning, and Equity in Education

Can We Have It All?

Clara Delavallade

Alan Griffith

Rebecca Thornton

Markets, Trade and Institutions Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

AUTHORS

Clara Delavallade (c.delavallade@cgiar.org) is a research fellow in the Markets, Trade an Institutions Division of the International Food Policy Research Institute, Washington, DC.

Alan Griffith (alangrif@umich.edu) is a graduate student in the Department of Economics, University of Michigan, Ann Arbor, MI, US.

Rebecca Thornton (rebeccat@illinois.edu) is an associate professor in the Department of Economics at the University of Illinois at Urbana-Champaign, Urbana, IL, US.

Notices

Contents

Tables

ABSTRACT

The United Nations Sustainable Development Goals have set a triple educational objective: improving access to, quality of, and gender equity in education. This study is the first to document the effectiveness of policies targeting all these objectives simultaneously. We examine the impact of a multifaceted educational program—delivered to 230 randomly selected primary schools in rural India—on students' participation and performance. We also study the heterogeneity of this impact across gender and initial school performance, and its sustainability over two years. Although the program specifically targeted out-of-school girls for enrollment, the learning component of the program targeted boys and girls equally. We find that the program reduced gender gaps in school retention and improved learning during the first year of implementation. However, targeting different educational goals (access, quality, and equity) did not yield sustained effects on school attendance or learning, nor did it bridge gender inequalities in school performance over the two-year period.

Keywords: development economics, education, gender, primary school enrollment, retention, student academic performance, India

ACKNOWLEDGMENTS

The authors thank the members of the NGO field team for their cooperation and contributions in all stages of the project and would especially like to acknowledge the useful comments provided by many seminar participants. Oxana Azgaldova and Flor Paz provided valuable research assistance. All errors are our own.

1. INTRODUCTION

The United Nations' Sustainable Development Goal 4 proposes that by 2030 "all girls and boys [should] complete free, equitable, and quality primary and secondary education" (United Nations 2015, 21). This goal sets a triple objective for educational policies: improving free access to, equity in, and quality of learning. Despite dramatic progress toward achieving higher school enrollment—with net enrollment rates in primary school reaching 88.5 percent in 2014 in low- and middle-income countries—gender inequalities in school participation remain high (UNESCO Institute of Statistics 2016). The situation in India is particularly stark, and gender gaps in access to education are among the largest in rural Rajasthan, with 8.9 percent of girls not in school, compared with 4.2 percent of boys (Pratham Organization 2011).[1] In this context, programs targeting girls' school enrollment and retention are critical for bridging gender gaps in access to education.

However, policies that promote access to school may not necessarily foster quality education. After much progress in improving school access, organizations and governments have shifted their focus toward improving learning. This second wave of programming has brought in a wide range of interventions, such as teacher training, performance incentives, and classroom inputs, and there now exists a significant amount of rigorous evidence on the impact of these programs on learning. Recent meta-analyses suggest moderate returns from isolated interventions, with gains in learning ranging between 0.04 and 0.15 standard deviations across 77 studies (McEwan 2015).[2] These analyses also suggest, based on factorial randomized designs, that some learning-oriented interventions, such as the use of instructional materials, may be effective only when combined with other interventions, such as providing teacher training (Glewwe, Kremer, and Moulin 2009; Banerjee et al. 2007; Glewwe et al. 2004), reducing class size and using contract teachers (Duflo, Dupas, and Kremer 2015; Muralidharan and Sundararaman 2010), or tracking students by ability (Duflo, Dupas, and Kremer 2011). What is less clear is how programs combining all three targets—enrolling marginalized students, increasing learning, and fostering gender equity in education—perform, especially over time. While the limited evidence on two-year educational programs has shown positive effects (Banerjee et al. 2007), incentives may not be aligned to allow for impact persistence over time.

One of the main challenges to successful implementation and evaluation is that meeting the first target—enrolling marginalized students—affects both the number and the type of students available to be targeted for the second objective—improving learning. When marginalized students are enrolled, the average learning level decreases and the variance in learning levels increases; in addition, larger class sizes put additional strain on teachers and resources. Each of these factors is individually associated with lower student outcomes, threatening program success and potentially adding downward bias to impact estimates based on aggregate-level data (Angrist and Lavy 1999; Duflo, Dupas, and Kremer 2011, 2015). Increasing the number of marginalized students may also increase students' propensity to drop out, increasing the potential for bias due to selection. Targeting female or marginal students may affect motivation among students who are not targeted, such as boys or high-performing students, suggesting the importance of taking heterogeneous treatment effects into consideration (Delavallade, Griffith, and Thornton 2016).

[1] Gender gaps are wider among marginalized populations. In South Asia, more than 40 percent of girls 15–19 years old from poor households have never completed first grade (Filmer and Pritchett 1999).

[2] For more experimental evidence, see Glewwe, Kremer, and Moulin (2009), who found no impact of textbook provision on average learning in Kenya, despite a significant positive impact on initially well-performing students. Glewwe and colleagues (2004) also found no impact of instructional materials in the form of flip charts on test scores. Duflo, Dupas, and Kremer (2011) found a positive impact of class size reduction when implemented in tandem with student tracking, especially for low-achieving students. Duflo, Hanna, and Ryan (2012) also found a positive impact of class size reduction only when combined with contract teachers. Banerjee and others (2007) found a positive impact of contract teachers in India, especially for low-performing students. Evidence on the impact of local governance programs is provided in Bold and colleagues (2013) and Duflo, Dupas, and Kremer (2015).

This paper examines the effects of a multifaceted educational program in Rajasthan, India, aimed at enrolling marginalized girls and improving student learning in primary school. Educate Girls, an India-based nongovernmental organization (NGO), implements a number of education interventions to increase student learning and reduce the gender gap in education in rural Rajasthan. The program studied in this paper consists of (1) enrolling marginalized girls who either never enrolled in school or dropped out, (2) training volunteers in child-centered techniques emphasizing activity-based and playful learning that groups students by ability, and (3) working with school management committees and members of the community to promote girls' education.

The literature suggests that although the school-participation and gender-equity components of the program—that is, girls' encouragement campaigns—have shown limited effects on learning compared with methods such as merit scholarships and school health programs (Kremer and Holla 2009), the learning components of the program—matching teaching to students' learning levels, improving accountability, and hiring local volunteers—have been shown to be among the most cost-effective ways of increasing learning (Kremer, Brannen, and Glennerster 2013). In addition, in a developing-country context, where most well-identified studies have failed to find an effect on test scores from class size reduction alone (Duflo, Dupas, and Kremer 2015) or improved pedagogy alone (Lucas et al. 2014), complementing these efforts with specifically trained contract teachers, pedagogy improvement, and governance-building strategies may substantially improve test scores.

We use a cluster-randomized experiment in 230 primary schools to examine the impact of the threefold program on students' participation and performance, the heterogeneity of the impact across gender and initial performance, and the sustainability of the impact. Using longitudinal individual-level data on enrollment, retention, attendance, and test scores in various subjects over two years, as well as the randomized design, we are able to address these questions without the usual causal concerns that come from analysis with cross-sectional or aggregate data. To minimize selection bias, we present results separately for newly enrolled students, and we discuss threats that may arise from differential dropout.

We find that overall, the program did not improve sustained school participation or gender equity in access to education. At the end of the two years of the program, there was no significant effect on the number, gender, and initial learning level of students enrolling in school. We find positive effects on retention. However, these effects dwindled over the course of the school year and did not translate into better attendance at the end of the year in program schools. Impacts on test scores at the end of the first year were high: 0.27 standard deviations on average, which is comparable to the impact of contract teacher provision programs (0.28 standard deviations in Banerjee et al. 2007; 0.14 standard deviations in Muralidharan and Sundararaman 2010; 0.21 standard deviations in Duflo, Dupas, and Kremer 2011).

Among newly enrolled children, girls benefited more than boys from the learning component, while among those previously enrolled, girls benefited less than boys. However, students at the bottom of the test score distribution did not experience these benefits. Although the distribution of post-program test scores in schools participating in the program exhibits first-order stochastic dominance relative to control school test scores—that is, beneficiary schools performed better at every point in the distribution—gains are not significant at the bottom and increase as one moves toward the high end of the score distribution. There are no statistically significant gains in learning due to the program after one year.

In other words, although the multifaceted program was successful at reducing the gender gap in school retention, combining interventions with different educational goals did not yield sustained effects on either school attendance or quality of learning. This combined program did not prove to be an effective strategy for bridging gender inequalities in school performance.

A large number of studies have rigorously evaluated the impact of educational programs in developing countries, including India. Randomized experiments have looked into the effectiveness of programs aimed at increasing school participation (Banerjee et al. 2007; Bobonis, Miguel, and Puri-Sharma 2006; Duflo, Hanna, and Ryan 2012; Muralidharan and Prakash 2013) and improving teaching quality (Banerjee et al. 2007; Borkum, He, and Linden 2012; Das et al. 2013; Muralidharan and Sundararaman 2010). Though most learning-targeted interventions have shown significant impacts on learning (Banerjee et al. 2007; He, Linden, and MacLeod 2008; Linden 2008), enrollment-targeted interventions have shown zero or small learning effects (Miguel and Kremer 2004; Petrosino et al. 2012). For instance, conditional cash transfers have been shown to improve school enrollment and educational attainment (Galiani and McEwan 2013) but to have limited effects on learning (Behrman, Parker, and Todd 2009). Taken together, these findings emphasize the complementarities between the two educational objectives but also the scarcity of evidence of these complementarities. Our study is one of the first to evaluate interventions combining enrollment and learning targets at the same time.

There is a growing body of evidence on the effectiveness of girl-focused interventions to address the gender gap in schooling, especially on programs that reduce direct costs (Bruns, Mingat, and Rakatomalala 2003; Deininger 2003) and indirect and opportunity costs (Khandker, Pitt, and Fuwa 2003; Lavinas 2001) and programs aimed at involving communities (Herz 2002; Benveniste and McEwan 2000), making schools girl-friendly (World Bank 2001; Herz 2002), and improving the quality of education (Lloyd, Mensch, and Clark 1998; Khandker 1996). However, to our knowledge, none of the previous studies have specifically explored the effect of encouragement on girls' schooling participation.

This paper also adds to the scant and inconsistent evidence regarding the impact of gender-targeted schooling programs on boys' outcomes. Kremer, Miguel, and Thornton (2009) found that boys benefited from a merit-based scholarship program even if they were not eligible themselves. In contrast, boys responded negatively to being excluded from a gender-based life skills program (Delavallade, Griffith, and Thornton 2016). We also contribute to recent research on "nudging" as a way to foster school participation. Benhassine and colleagues (2013) suggest that small "nudges" may be sufficient to significantly increase human capital investment, based on the finding that simply labeling cash transfers as an education support program caused large gains in school participation. The enrollment drive aspect of the program we study relied only on encouragement through community sensitization about girls' education and door-to-door visits by volunteers, and showed no significant impact on school participation. This finding suggests that encouragement needs to be accompanied by economic incentives, even if small ones, for nudging to affect school participation. Finally, our paper adds to the literature on the sustainability of treatment effects over time (Banerjee et al. 2007; Kremer and Miguel 2007; Andrabi, Das, and Khwaja 2008; Kremer, Miguel, and Thornton 2009; Duflo, Dupas, and Kremer 2011; Baird et al. 2016).

The remainder of the paper is organized as follows. Section 2 presents the study context and describes the multifaceted education intervention. Section 3 describes the experimental design and outcome measures. Section 4 presents the data and empirical strategy. Sections 5 and 6 discuss methodological challenges and present the results, respectively. Section 7 further discusses some of our results, and Section 8 concludes.

2. BACKGROUND

Schooling in Rajasthan: Participation, Quality, and Gender Equity

Despite educational advances in most developing countries, the state of Rajasthan in India has experienced limited educational gains over the past decade, especially for girls. In Rajasthan, girls continue to face strong barriers to education that can be attributed to gender (World Bank 2011). In 2012, 4.6 percent of girls 7–10 years old were still not in school in rural Rajasthan, compared with 2.2 percent of boys (Pratham Organization 2012). This gender gap widens considerably as students age, due largely to social norms that are particularly discriminatory toward girls. In rural India, marriage is seen as a substitute for schooling for many girls, and girls often have no say in when and whom they marry. In addition, marriage often occurs at a young age, with 57.6 percent of women marrying younger than the legal age (UNICEF 2012, 173). Girls 11–14 years old are the hardest to keep in school. In Rajasthan, the proportion of out-of-school girls in this age range increased from 8.9 percent in 2011 to more than 11 percent in 2012 (Pratham Organization 2012).

In addition, educational quality is low, with only 47.7 percent of children in grades three to five able to read a grade one–level text in government schools in 2011, and only 33.1 percent able to do subtraction in 2012 (Pratham Organization 2012). The availability of primary schools in remote areas is still limited, leading to high variance in student-teacher ratios. In addition, teacher training has not yet been unified in India, leading to large discrepancies in teacher quality.

The Educate Girls Intervention

We evaluate an intervention developed and implemented by Educate Girls, an Indian NGO working with government schools in the state of Rajasthan. One of its main program aims is to increase girls' educational outcomes, with a focus on children in lower primary school (grades one through five).

The program consists of several components that separately target enrollment, quality education, and school management and is directed by a trained volunteer in each village.

First, Educate Girls conducts extensive efforts to enroll and retain girls in school. Before each school year, it identifies out-of-school girls using information from community members and government records. The volunteers hold village meetings to prepare for a house-to-house enrollment drive, targeting girls who have never enrolled or who have dropped out of school. These efforts seek both to encourage parents to support their daughters' education and to motivate girls themselves to come to school.

The second component of the program targets student learning and involves in-school lessons led by the volunteer. The Creative Learning and Teaching (CLT) curriculum was designed with Pratham Rajasthan and emphasizes activity-based and playful learning through games that teach English, Hindi, and math. CLT relies on two teaching approaches to account for different learning levels within the classrooms. First, the "peer group learning" methodology emphasizes group work and student involvement in the teaching and learning process. This approach has the combined effect of making the learning process more engaging and fun for students while decreasing the burden placed on teachers.

In tandem with the peer group learning method, teachers also use a "catch-up" methodology to close the learning gap for students lagging behind the rest of the class. Prior to implementing the new curriculum, CLT-trained teachers or Educate Girls volunteers conduct a group of diagnostic pre-program tests based on the Annual Survey of Education Report (ASER) standards to identify children who are struggling to master basic concepts. The tests are designed to be quick to administer so that they can be conducted individually for each student in the program. In schools where the number of children scoring poorly is greater than 60, an Educate Girls staff member provides tailored "hand-holding support" to assist teachers.

This program does not focus explicitly on girls, but rather aims to increase learning levels for both girls and boys in grades three through five. These additional lessons are held during school hours for approximately two hours per day, several days per week. They are conducted over four to five months.

The third component of the intervention involves working with school management committees (SMCs) and community volunteers in each village. SMC members receive training and support from the Educate Girls volunteer and staff to build capacity, increase parent engagement, formulate and follow an annual school improvement plan (SIP), and sensitize the community to issues related to girls' education. This program activity is implemented throughout the academic year.

Educate Girls began implementation of the program in 2011. However, programmatic difficulties resulted in weak program fidelity. We evaluate the program for the academic years of 2012 and 2013, when the program was fully and correctly implemented.

3. RESEARCH DESIGN

Sample and Randomization

The study consists of 230 primary schools located in 98 villages in Rajasthan. Villages with at least 1 government primary school in four administrative blocks[3] were selected to be included in the evaluation.

We study the effects of the program on two distinct samples of students. Our *Baseline Sample* consists of students who were enrolled at the beginning of 2011 in grades three and four. Our *Newly Enrolled Sample* consists of students who were not enrolled at the baseline in 2011 but enrolled in grades three, four, or five in either 2012 or 2013. Students were identified using enrollment registers collected each year. These enrollment registers also listed each student's gender and age, as well as whether he or she belonged to a Scheduled Caste, Scheduled Tribe, or Other Backward Caste.

Prior to implementation of the program, the researchers randomly assigned villages to either treatment or control groups, stratified by school size.

Student-Level Data

Students in the samples took a standard test twice in each academic year, 2012 and 2013. The pre-program test took place prior to program implementation in a given year, while the post-program test was administered after completion of the learning component of the program during that year. Using all rounds of test score data (pre- and post-program tests in 2012 and 2013), we measure attendance by whether the student was present on the day that tests were administered. We focus on treatment effects on attendance using data collected after program implementation.

Conditional on the student being present in school on the day of the test, the test measured knowledge in three subjects—Hindi, English, and math.[4] The tests were scored categorically from A to E, with A being the highest score and E the lowest. We assigned each letter grade a numeric value from 1 to 5 and normalized by subtracting the mean score and dividing by the standard deviation of the control. We evaluate learning outcomes using the normalized test scores from the post-program exams.

To match test score and attendance data in 2012 and 2013 with the baseline 2011 enrollment list, we used information regarding each student's school, village, gender, age, last and first name, and father's name. In cases in which there were multiple possible matches, we kept the observations in the data but weighted all analyses by the inverse of the number of possible matches. Of students enrolled at baseline and in appropriate grades, 79.5 percent were successfully matched to the 2012 test score data, while 61.7 percent were successfully matched to the 2013 test score data.[5]

We collected data for students in grades three through five in each respective academic year. Baseline data are only available for students in grades three through five in 2011. This implies that newly enrolled students in grade three in 2012 as well as newly enrolled students in grades three and four in 2013 do not have baseline enrollment data.

School-Level Data

In addition to individual-level data, we collected school-level information that we use in the analysis. We use data from India's District Information System for Education (DISE), collected prior to the introduction of the program in 2011. We supplemented these data with data including measures of school infrastructure (such as the presence of electricity and computers and the number of students at each school) collected by Educate Girls staff. We used these data to check for baseline balance and as covariates in the analyses.

[3] Ahore, Jalore, Jaswantpura, and Sayla.

[4] See www.asercentre.org for more information about the ASER tool.

[5] The matching success rate did not significantly differ across treatment and control students. Results are available upon request.

We used school rosters collected in 2011, 2012, and 2013 to construct grade-level enrollment numbers. We also collected school-level SMC outcomes, including the number of SMC meetings held and the number of SIPs prepared and implemented. We collected these outcomes monthly from July 2012 to January 2013, for a total of seven observations per school.

4. EMPIRICAL STRATEGY

Identification of Effects

To measure the main treatment effects of the program on student-level school participation outcomes, we estimate the specification

$$Y_{ijt} = \beta_0 + \beta_1 T_j + \epsilon_{ijt} \tag{1}$$

for student *i* in village *j* at time *t*. T_j indicates whether village *j* was assigned to the treatment group. The main dependent variables, denoted as Y_{ijt}, include participation outcomes measured in 2012 and 2013: an indicator for being enrolled in school at time *t* and indicators for being present in school on given days. We estimate the equation with a linear probability model and cluster standard errors by village, the unit of randomization. As described above, we weight student observations by the inverse of the number of potential matches between 2011 enrollment and test score data collected in subsequent years.

We also measure treatment effects on learning. We estimate the following specification:

$$Y_{ijst} = \beta_0 + \beta_1 T_j + \epsilon_{ijst}, \tag{2}$$

where variable Y_{ijst} is learning outcome for individual *i* in village *j* on test subject *s* at time *t.* We present results for two learning outcomes. First, we present results for normalized test scores. These scores are created by converting the raw test scores (scored A, B, etc.) into discrete outcomes (5, 4, etc.) and then normalizing. Second, we present results where the learning outcome is an indicator for test score improvement. For purposes of defining improvement, we compare test scores administered prior to program implementation in 2012 to test scores taken at three subsequent administrations: 2012 post-test, 2013 pre-test, and 2013 post-test. We include subject (English, Hindi, and math) and grade (three or four) indicators in some specifications.

Additionally, we test for heterogeneous effects along two dimensions: gender and baseline ability. To test for differential effects by gender, we also estimate Equations (3) and (4), which are analogous to Equations (1) and (2), respectively

$$Y_{ijt} = \beta_0 + \beta_1 T_j + \beta_2 Girl_{ij} + \beta_3 T_j \times Girl_{ij} + \epsilon_{ijt} \tag{3}$$

and

$$Y_{ijst} = \beta_0 + \beta_1 T_j + \beta_2 Girl_{ij} + \beta_3 T_j \times Girl_{ij} + \epsilon_{ijst}. \tag{4}$$

Inasmuch as the program explicitly targets girls, we are interested in the size and statistical significance of the coefficient β_3 in Equations (3) and (4). Because the program focuses explicitly on girls' enrollment, we would expect this coefficient to be positive when estimating effects on enrollment and retention, indicating a larger treatment effect on girls than on boys. In contrast, because all students—both boys and girls—received the in-class learning component of the program, there is unlikely to be an additional effect on girls' performance, unless girls respond better than boys to the program's type of teaching.

We are also interested in testing for differential effects of the program by pre-program ability. In some specifications, we condition on 2012 pre-program test scores when available (indicators of baseline grade levels, A through E), to see whether treatment effects vary based on students' pre-program ability.

All our results are presented separately for students in the Baseline Sample or Newly Enrolled Sample. Because one of the program's explicit goals is to bring in never-enrolled and dropout students (with an emphasis on girls) during the enrollment drive, we might expect new enrollees to be different—in number and type—across treatment and control schools in the Newly Enrolled Sample. We discuss the potential bias due to differential selection below.

Results on attendance are only observed conditional on enrollment. Similarly, test score results are only observed conditional on attendance on the day of the exam. Lastly, analysis using pre-program test scores relies on a student having taken the pre-program test. We discuss potential bias due to differential enrollment, retention, and attendance below.

In addition to the individual-level analyses, we estimate treatment effects on class-level enrollment and school-level SMC outcomes as described above.

Baseline Balance

An important identification assumption for our empirical strategy to attribute causal effects to the Educate Girls program is that there are no systematic differences between observations in the treatment and control schools at baseline. Table 4.1 presents balancing tests of student- and school-level characteristics from 2011, prior to program implementation. The Baseline Sample comprises a total of 7,309 students, equally split between treatment and control.[6] Panel A, Column 4, presents the *p*-values from testing the equality of means of baseline variables across the treatment and control groups. We see no significant differences between treatment and control students across gender, caste, type of school (upper or lower secondary), or school characteristics such as having electricity, a computer, or drinking water. The magnitudes of the differences between treatment and control groups, presented in Column 3, are also all relatively small. Further, we see no differences in test scores or school attendance between treatment and control groups prior to program implementation in 2012 (Panel B, Column 4).

Table 4.1 Pre-program characteristics of students: Baseline sample

Variable	Treatment	Control	Difference	*P*-value of balancing test
Panel A: Baseline characteristics across treatment				
Girl	0.526	0.547	-0.021	0.273
	(0.014)	(0.014)	(0.019)	
Grade 4 at baseline	0.461	0.461	0.000	0.993
	(0.011)	(0.009)	(0.014)	
Scheduled Caste	0.295	0.300	-0.005	0.903
	(0.025)	(0.030)	(0.039)	
Scheduled Tribe	0.177	0.155	0.022	0.568
	(0.032)	(0.022)	(0.039)	
Other Backward Caste	0.370	0.406	-0.035	0.491
	(0.037)	(0.035)	(0.051)	
Upper primary school	0.560	0.588	-0.027	0.761
	(0.054)	(0.073)	(0.090)	
Number of students enrolled in school	60.271	64.481	-4.210	0.518
	(4.107)	(5.070)	(6.491)	
Electricity	0.400	0.384	0.016	0.861
	(0.069)	(0.064)	(0.094)	
Computer	0.086	0.104	-0.018	0.714
	(0.034)	(0.037)	(0.050)	

[6] There are 7,309 students in the Baseline Sample after matching weights are applied. The gross number of students in the sample is 7,782 but includes multiple entries of some observations due to the matching process. See the appendix for details on matching and weighting.

Table 4.1 Continued

Variable	Treatment	Control	Difference	*P*-value of balancing test
Panel A: Baseline characteristics across treatment (continued)				
Drinking water	0.818	0.808	0.010	0.898
	(0.050)	(0.061)	(0.078)	
Teacher-pupil ratio	0.067	0.070	-0.002	0.793
	(0.005)	(0.007)	(0.008)	
Panel B: Outcomes measured before program exposure				
Attended 2012 pre-program test	0.558	0.544	0.014	0.591
	(0.018)	(0.020)	(0.027)	
Score on 2012 pre-program test (Hindi)	3.255	3.310	-0.055	0.574
	(0.075)	(0.061)	(0.097)	
Score on 2012 pre-program test (English)	2.374	2.418	-0.044	0.677
	(0.083)	(0.065)	(0.105)	
Score on 2012 pre-program test (math)	3.061	3.119	-0.057	0.603
	(0.076)	(0.080)	(0.110)	

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. The *p*-value is the result of the difference in means between treatment and control. The null hypothesis is that they are equal. The sample is students who were enrolled in grades 3 or 4 in 2011. $N = 7{,}309$ students in 230 schools. All analyses are performed using weights to account for multiple matching.

5. METHODOLOGICAL CHALLENGES: DIFFERENTIAL SELECTION AND ATTRITION

Before turning to the results, we first discuss potential threats to the validity of our analysis due to selection and attrition.

Potential Threats

The data we use to measure treatment effects on attendance and test scores were collected at schools. Thus, if there were differential enrollment or retention across treatment and control schools, the effect of the program on later-measured outcomes is likely to be biased. Estimates of the program's impact on test scores may be biased downward if the program is successful at targeting low-performing students. Conversely, these estimates may be biased upward if high-performing students are less likely to drop out or be absent on the test day in treatment schools. The Educate Girls program's main components aimed at enrolling marginalized students—specifically targeting girls prior to the beginning of the school year, retaining students, and improving learning. These components provide two potential sources of selection bias that may threaten the validity of our analysis of the program's effect on students' attendance and performance.

1. *Differential selection*: If Educate Girls is successful at enrolling marginalized students, newly enrolled students may have different characteristics in treatment schools than in control schools.
2. *Differential attrition*: Over the course of the school year, students in the control schools may be more likely to drop out, and those students might have specific characteristics (for example, they are lower performers) that would lead to students in the treatment and control schools being significantly different at the end of the school year.

A selection bias arises if newly enrolled students' and/or attriters' characteristics are indeed significantly different between the treatment and control groups *and* if those characteristics are correlated with the outcomes of interest (school attendance or test scores).

Recall that the Baseline Sample is comprised of students enrolled at baseline at the beginning of the 2011 school year. Accordingly, this sample can suffer only from differential attrition. The Newly Enrolled Sample, in contrast, is composed of students who enroll right after the first or second enrollment drive in 2012 or 2013. Differential selection and attrition may pose threats to the validity of estimates on this sample. We provide quantitative evidence on these potential sources of bias below.

Evidence

Differential Selection

We found no significant differences between students newly enrolled in treatment and control schools in terms of observable characteristics such as gender, school type, grade, initial attendance, and academic performance (Table 5.1). This indicates a low risk of selection bias due to differential enrollment, although we cannot entirely rule out selection bias due to differences in unobservable characteristics of newly enrolled students in treatment and control schools.

Table 5.1 Pre-program characteristics of students: Newly enrolled sample

Variable	Treatment	Control	Difference	*P*-value of balancing test
Panel A: Baseline characteristics across treatment				
Girls	0.537	0.486	0.050	0.188
	(0.030)	(0.026)	(0.038)	
Grade 4 at baseline	0.279	0.251	0.028	0.338
	(0.022)	(0.024)	(0.030)	
Upper primary school	0.676	0.600	0.076	0.522
	(0.065)	(0.101)	(0.119)	
Number of students enrolled in school	63.793	64.397	-0.604	0.940
	(5.439)	(5.938)	(7.999)	
Electricity	0.518	0.449	0.070	0.546
	(0.078)	(0.090)	(0.115)	
Computer	0.105	0.151	-0.046	0.599
	(0.057)	(0.068)	(0.088)	
Drinking water	0.879	0.830	0.049	0.470
	(0.036)	(0.058)	(0.067)	
Teacher-pupil ratio	0.076	0.069	0.007	0.503
	(0.007)	(0.008)	(0.011)	
Panel B: Outcomes measured before program exposure				
Attended 2012 pre-program test	0.382	0.466	-0.083	0.060
	(0.032)	(0.037)	(0.044)	
Score on 2012 pre-program test (Hindi)	3.315	3.506	-0.191	0.158
	(0.106)	(0.083)	(0.134)	
Score on 2012 pre-program test (English)	2.526	2.553	-0.027	0.870
	(0.120)	(0.111)	(0.163)	
Score on 2012 pre-program test (math)	3.180	3.262	-0.082	0.625
	(0.124)	(0.114)	(0.167)	

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. The Newly Enrolled Sample is students who would have been in grades 3 or 4 at baseline, but were not enrolled at baseline and enrolled in either 2012 or 2013. That is, they first enrolled in grades 4 or 5 in 2012, or grade 5 in 2013. N = 1,058 students in 230 schools. All analyses are performed using weights to account for multiple matching.

Differential Attrition

Had the program successfully retained lower-performing students in treatment schools, estimates of the program's impact on learning might be plagued by selection bias due to differential attrition in both groups. However, the absence of evidence of differential attrition by pre-program test score for either sample mitigates this concern (Table 5.2). In Columns 5 and 8, we interact a treatment indicator with 2012 average pre-program test scores,[7] showing that there is no significant difference in retention based on this score for students enrolled at baseline. We see similar evidence for the Newly Enrolled Sample, with no significant difference in retention until 2013 (year two of implementation) based on the students'

[7] Since tests were taken on three subjects (Hindi, English, and math), we use the average of the three scores as a proxy for baseline ability.

test scores on the 2012 pre-program test. We see a similar lack of evidence of differential attrition as defined by differential attendance at school on the dates that tests were administered.[8]

Table 5.2 Treatment effects on retention of baseline and newly enrolled students

	Baseline sample					Newly enrolled sample		
	Enrolled in 2012		Enrolled in 2013			Enrolled in 2013		
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
T	0.045**	0.030	0.069**	0.046	0.098	0.007	0.047	0.098
	(0.022)	(0.027)	(0.033)	(0.038)	(0.088)	(0.042)	(0.050)	(0.257)
Girl		0.025		-0.013			0.035	
		(0.019)		(0.036)			(0.032)	
T × Girl		0.030		0.042			-0.080*	
		(0.025)		(0.043)			(0.042)	
Mean score on 2012 pre-test					0.018			0.019
					(0.023)			(0.065)
T × Mean score on 2012 pre-test					-0.008			-0.027
					(0.028)			(0.080)
Constant	0.773***	0.768***	0.583***	0.590***	0.732***	0.830***	0.813***	0.698***
	(0.016)	(0.020)	(0.022)	(0.027)	(0.073)	(0.030)	(0.038)	(0.216)
Cohort controls	NO	YES	NO	NO	NO	NO	NO	NO
Observations	7,782	7,782	4,333	4,333	2,406	744	744	244
R-squared	0.003	0.007	0.005	0.006	0.011	0.000	0.003	0.001
Mean of dep. var. in control	0.773	0.773	0.583	0.583	0.583	0.830	0.830	0.830
P-value: T + T × Girl = 0		0.012		0.030	0.153		0.456	0.693

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). The samples in Panel A, Column 5, and Panel B, Column 3, are conditional on having taken the 2012 pre-test. All analyses are performed using weights to account for multiple matching.

[8] For the sake of brevity, we have not included these results here. They are available upon request from the authors.

6. RESULTS

Here, we present reduced-form estimates of the effect of the intervention on enrollment and retention, attendance, and test scores, as well as on SMC outcomes. For each outcome, we explore differential effects by gender and baseline ability. We present the results on outcomes in 2012 and 2013.

Enrollment and Retention

We first present the effects of the program on enrollment, using school-level and individual-level data. We then turn to examining the program's effects on retention.

Total Enrollment

We first examine the effects of the program on total enrollment at the school level using data from classroom rosters.[9] Table 6.1 shows estimates of the program's impact on the number of students enrolled. It presents results for 2012 and 2013, for all students and for girls only. There is no significant difference in the number of students enrolled (newly enrolled or not) between treatment and control schools in either year. Panel B presents analogous results for the percentage change in students enrolled since 2011, showing that the program had no impact on percentage change in enrollment since the program was introduced.

Table 6.1 Treatment effects on grade-level total enrollment

Variable	Treatment	Control	Difference	*P*-value of balancing test
Panel A: Number of students enrolled by grade				
Total enrolled in 2012	15.935	15.638	0.297	0.815
	(0.597)	(0.581)	(1.264)	
Number of girls enrolled in 2012	8.561	8.534	0.027	0.970
	(0.348)	(0.337)	(0.717)	
Total enrolled in 2013	15.232	14.561	0.671	0.596
	(0.588)	(0.563)	(1.264)	
Number of girls enrolled in 2013	8.172	7.799	0.373	0.600
	(0.329)	(0.339)	(0.711)	
	Treatment	**Control**	**Difference**	***P*-value of balancing test**
Panel B: Percent changes in number of students enrolled by grade				
Percent change in total enrolled in 2012	0.202	0.186	0.016	0.783
	(0.037)	(0.046)	(0.057)	
Percent change in number of girls enrolled in 2012	0.340	0.295	0.045	0.509
	(0.057)	(0.057)	(0.067)	
Percent change in total enrolled in 2013	0.195	0.114	0.081	0.252
	(0.044)	(0.042)	(0.071)	
Percent change in number of girls enrolled in 2013	0.392	0.224	0.168	0.243
	(0.094)	(0.069)	(0.143)	

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. The sample is school-by-grade aggregates, comprised of 3 grades each in 230 schools (total sample of 690).

[9] There was no significant difference between treatment and control schools in the number of students enrolled at baseline (not shown here).

Characteristics of New Enrollees

In addition to finding no effect on the total number of students enrolled, we further find no evidence that the program affected the type of students enrolled, despite Educate Girls' targeting efforts. Table 5.1 presents the characteristics of the Newly Enrolled Sample, consisting of students in grades three or four who first enrolled in the sample schools in 2012 or 2013. In total, there were 1,058 new students over these two years. Although Educate Girls targeted its efforts at enrolling more girls, new enrollees were only 5 percentage points more likely to be girls in treatment schools than in control schools, a difference that is not statistically significant (Panel A, *p*-value = 0.188). New enrollees did not differ between treatment and control schools in terms of observable grade and school characteristics, such as having a computer or electricity. This finding suggests that the enrollment drive did not lead to a differential selection of students based on those characteristics. Using 2012 pre-program test score data, we find no significant differences between newly enrolled students in treatment schools and those in control schools for 2012 pre-program test attendance or for any of the 2012 pre-program test scores (Panel B).

Retention

We turn next to measuring the effects of the program on retention. Recall that the program's enrollment drive specifically targeted dropout or never-enrolled girls. While the program did not actively aim to keep currently enrolled girls in school, the program's community sensitization component may have affected the retention of girls already enrolled at baseline.

In contrast to the lack of treatment effects on new enrollments, we find evidence that the program was successful in *retaining* girls and lower-performing students in school during subsequent years. Columns 1 to 5 in Table 5.2 show the results for the Baseline Sample. Overall retention of those who were enrolled in 2011 to 2012 in the control schools was 77.3 percent. Column 1 indicates that students in treatment schools were 4.5 percentage points more likely to be enrolled one year later, a 5.8 percent increase in enrollment compared with control schools. This effect is statistically significant (*p*-value = 0.040) and concentrated on girls in treatment schools, who were 6 percentage points more likely to still be enrolled in school in 2012 (*p*-value = 0.012) than girls in control schools, as seen in Column 2.

The retention effects among Baseline Sample students are similarly high in 2013. The main treatment effect is 6.9 percentage points, or a 12.3 percent increase in the likelihood of being retained over those in the control (Column 3). Again, this effect is higher for girls in treatment schools, who were 8.8 percentage points more likely to re-enroll in school than girls in control schools (*p*-value = 0.030). These results indicate that the Educate Girls program resulted in higher rates of school retention over the first two years of implementation, especially among girls.

Columns 6 to 8 present the effects of the program on retention of students in the Newly Enrolled Sample. Overall, retention of newly enrolled students in control schools from 2012 to 2013 was 83 percent. We find no statistically significant effects of the program on retention after one or two years, and the magnitude of the difference in retention is small. Note also that there was no differential effect for girls, despite the gender-specific enrollment targeting (*p*-value = 0.456). The interaction between treatment and being a girl is negative, going in the opposite direction of girl-specific targeting. Results in Columns 6 to 8 show an absence of impact on retaining students in the Newly Enrolled Sample.

Table A.1 in the appendix investigates the heterogeneity of the effects on retention from 2012 to 2013, by 2012 pre-program test score.[10] From this, we see that for the Baseline Sample the effect is positive throughout the distribution and largest among those with a baseline score of A, the highest score. Further, the significantly higher treatment effect on retention for girls noted above is driven by girls with the highest pre-program test scores.

[10] Note that in the specifications for Table A.1, there are three observations per individual, corresponding to the three subjects—Hindi, math, and English.

Attendance

Table 6.2 presents the results on the effect of the program on attendance in 2012 and 2013, as measured by whether the student was present on the day when tests were administered. Overall, the results on attendance are in line with the results on retention, with the program's only impact being on students in the Baseline Sample (Panel A) and the only significant effect on the 2013 pre-test.

We see no significant effect on attendance at the 2012 post-program test, as shown in Columns 1 and 2. In contrast, Column 3 shows a 6.8 percentage point increase in attendance at the 2013 pre-program test. This effect diminishes to 0.053 for the 2013 post-program test and is statistically insignificant (Table 6.2, Panel A, Columns 5 and 6).

Table A.2 in the appendix presents these results disaggregated by 2012 pre-test score. Similar to the retention effects, the treatment effects on attendance among the Baseline Sample are driven by higher-ability students, and especially higher-ability girls (Table A.2, Panel A) in the Baseline Sample.

Turning to students in the Newly Enrolled Sample (Table 6.2, Panel B), we find no statistically significant average difference in attendance overall. However, from Column 2, we see that the program has a statistically significant and large *negative* effect on boys' attendance at the 2012 post-program test, of 12.4 percentage points; this negative effect is not present for girls (p-value = 0.823).[11] The negative effect on boys' attendance disappears in 2013, or when conditioning on 2012 pre-program test scores (Table A.2, Panel B, Columns 1 and 2).

Taken together, these results suggest that Educate Girls was effective at retaining students on the enrollment list and at encouraging students to attend school at the beginning of the academic year, but less effective in motivating their attendance later in the academic year. Further, it appears that the program negatively affected *boys'* attendance at the end of the first year of program implementation but otherwise had no effect on attendance for newly enrolled students.

Table 6.2 Treatment effects on attendance on test days

Panel A: Baseline Sample Dependent variable: Present for test	2012 post-program test (1)	 (2)	2013 pre-program test (3)	 (4)	2013 post-program test (5)	 (6)
T	-0.013	-0.036	0.068**	0.063*	0.053	0.040
	(0.027)	(0.032)	(0.030)	(0.034)	(0.033)	(0.037)
Girl		-0.042**		-0.043		-0.053*
		(0.019)		(0.031)		(0.031)
T × Girl		0.041		0.009		0.023
		(0.027)		(0.041)		(0.040)
Constant	0.556***	0.581***	0.434***	0.457***	0.393***	0.423***
	(0.018)	(0.022)	(0.019)	(0.023)	(0.023)	(0.027)
Cohort controls:	NO	YES	NO	YES	NO	YES
Observations	7,782	7,782	4,333	4,333	4,333	4,333
R-squared	0.000	0.001	0.005	0.006	0.003	0.005
Mean of dep. var. in control	0.556	0.556	0.434	0.434	0.393	0.393
P-value: T + T × Girl = 0		0.845		0.062		0.124

[11] This result is similar to that found by Delavallade, Griffith, and Thornton (2016), who saw negative effects from a girls' parliament program on boys in India.

Table 6.2 Continued

Panel B: Newly Enrolled Sample **Dependent variable: Present for test**	**2012 post-program test**		**2013 pre-program test**		**2013 post-program test**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
T	-0.059	-0.124***	0.010	0.045	0.006	0.035
	(0.044)	(0.041)	(0.044)	(0.054)	(0.050)	(0.064)
Girl		-0.110***		-0.015		-0.039
		(0.029)		(0.040)		(0.059)
T × Girl		0.113***		-0.071		-0.059
		(0.042)		(0.053)		(0.077)
Constant	0.454***	0.423***	0.601***	0.609***	0.540***	0.559***
	(0.031)	(0.030)	(0.029)	(0.040)	(0.034)	(0.051)
Cohort controls:	NO	YES	NO	YES	NO	YES
Observations	1,079	1,079	744	744	744	744
R-squared	0.004	0.102	0.000	0.004	0.000	0.006
Mean of dep. var. in control	0.454	0.454	0.601	0.601	0.540	0.540
P-value: T + T × Girl = 0		0.823		0.596		0.706

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Attendance at each test administration is defined as the student having taken the test, such that a score was recorded. Cohort controls consist of indicators for being in the cohort in grade 4 at baseline (grade 3 is the omitted category). All analyses are performed using weights to account for multiple matching.

Learning

We now first present the effects of the program on student test scores immediately after being treated (that is, post-program tests in the same year as treatment). We then turn to examining potential persistent effects of the program in the years after being treated (that is, pre-program tests in the year after treatment). Test results are presented pooling all subjects—Hindi, English, and math.12

Immediate Treatment Effects

Table 6.3 presents the treatment effects on 2012 post-program test scores. The effects are positive, large, and significant for the Baseline Sample and the Newly Enrolled Sample. Treatment students scored 0.27 and 0.25 standard deviations higher than control students in each of the samples, respectively (Column 1). Comparing these effects with the effects on learning found in other randomized interventions, these gains are about twice as large as those of learning-targeted interventions focusing on one single educational input, such as using contract or volunteer teachers (average effect size = 0.10 standard deviations) or training teachers (average effect size = 0.12 standard deviations) (McEwan 2015). These gains are robust to including student- or school-level controls.

Further, we see positive impacts for both girls and boys in both subsets of students, and we are unable to reject that the effect is the same for boys and girls in both samples (Column 2), despite girls performing significantly lower than boys at the post-program test. This finding is not particularly surprising, given that the learning aspect of the program did not specifically target girls. Impact estimates disaggregated by subject indicate that the program benefits are mostly focused on Hindi and math, especially for girls, but are of a similar magnitude for English learning (Table A.3 in the appendix).

[12] Impact estimates display very similar patterns across subjects. See Table A.3 for the results for the 2012 post-test. Results for the 2013 pre- and post-program tests are available upon request.

In Table 6.3, Columns 3 and 4 present the results for outcomes defined as whether the 2012 post-program test score increased from the 2012 pre-program test score, after one program-year of exposure. Overall, students in both samples were more than 20 percentage points more likely to improve their score, with significant effects on both boys and girls.

Table 6.3 Treatment effects on 2012 post-program test scores (all subjects)

Panel A: Baseline sample	Normalized test score		Probability of improving from pre-program test	
	(1)	(2)	(3)	(4)
T	0.271***	0.321***	0.238***	0.234***
	(0.095)	(0.103)	(0.058)	(0.062)
Girl		-0.088**		0.003
		(0.043)		(0.022)
T × Girl		-0.095		0.007
		(0.069)		(0.028)
Constant	-0.000	0.331***	0.516***	0.574***
	(0.057)	(0.071)	(0.052)	(0.056)
Subject and cohort controls:	NO	YES	NO	YES
Observations	12,906	12,906	9,424	9,424
R-squared	0.017	0.192	0.061	0.071
Mean of dep. var. in control	0.000	0.000	0.516	0.516
P-value: T + T × Girl = 0		0.025		0.000
Panel B: Newly enrolled sample	**Normalized test score**		**Probability of improving from pre-program test**	
	(1)	(2)	(3)	(4)
T	0.254**	0.217	0.282***	0.369***
	(0.115)	(0.136)	(0.072)	(0.087)
Girl		-0.190**		0.121**
		(0.091)		(0.051)
T × Girl		0.069		-0.177**
		(0.137)		(0.075)
Constant	0.000	0.455***	0.427***	0.455***
	(0.087)	(0.089)	(0.058)	(0.074)
Subject and cohort controls:	NO	YES	NO	YES
Observations	1,386	1,386	998	998
R-squared	0.015	0.231	0.080	0.102
Mean of dep. var. in control	0.000	0.000	0.427	0.427
P-value: T + T × Girl = 0		0.030		0.011

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1 and 2 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 3 and 4 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). Cohort controls consist of indicators for being in the cohort in grade 4 at baseline (grade 3 is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Table 6.4 reports impact estimates disaggregated by 2012 pre-program test scores. We see that the positive effects on learning are concentrated among students in the middle of the distribution and increase as one moves upward in the score distribution, with the lowest-scoring students failing to experience the program's learning benefits.[13] While noting power issues due to the relatively small number of students scoring an E at baseline, we see no significant effects of the program on the normalized tests score for those who received the lowest score on the 2012 pre-program test.

We see similar results when examining the probability of improvement disaggregated by pre-program test score (Columns 6 to 9). Although the distribution of test scores in schools participating in the program exhibits first-order stochastic dominance relative to control school test scores—that is, beneficiary schools performed better at every point in the distribution—the gains appear most concentrated among those students who were already performing at the higher end of the test score distribution. There is no significant impact for the lowest pre-program test performers among those in either sample, and the effect increases in magnitude as students score higher on the pre-program test, among students in both samples. Students in the Baseline Sample scoring a D on the 2012 pre-program test were 34.8 percentage points more likely to improve their score (to A, B, or C) on the 2012 post-program test in treatment schools than in control schools (Panel A, Column 7). Those scoring a C on the 2012 pre-program test were 32.7 percentage points more likely to improve their score (to A or B) (Column 8), while those scoring a B on the 2012 pre-program test were 45.6 percentage points more likely to improve their score (to A) than students in control schools (Column 9). Similar patterns emerge for students in the Newly Enrolled Sample (Panel B).

Finally, we see no evidence that the treatment effect on test scores, disaggregated by pre-program test performance, is different for boys and girls in any specification, as indicated by the lack of significance of the estimated interaction term in even-numbered columns.[14]

[13] Learning impacts for best-performing students (those scoring an A) on pre-program tests are not significant. A significant effect here would be implausible, as it would require students in control schools to drop from the highest score at a higher rate than students in treatment schools.

[14] A slight exception to this pattern is for the 43 students in the Newly Enrolled Sample who scored an E on the 2012 pre-program test (Panel B, Column 2).

Table 6.4 Treatment effect on 2012 post-program test: By 2012 pre-program test (all subjects)

Panel A: Baseline sample	**Normalized test score**					**Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	0.046	0.166**	0.220***	0.379***	0.061	0.258	0.348***	0.327***	0.456***
	(0.084)	(0.071)	(0.079)	(0.068)	(0.040)	(0.173)	(0.084)	(0.090)	(0.089)
Girl	0.060	0.019	-0.025	-0.045	-0.065*	0.067	0.021	-0.034	-0.053
	(0.037)	(0.030)	(0.032)	(0.033)	(0.033)	(0.068)	(0.032)	(0.037)	(0.040)
T × Girl	-0.044	-0.013	0.018	0.039	0.030	-0.176	-0.052	0.035	0.035
	(0.078)	(0.039)	(0.043)	(0.048)	(0.052)	(0.153)	(0.048)	(0.054)	(0.061)
Constant	0.822***	0.691***	0.623***	0.341***	1.630***	-1.433***	-0.546***	0.333***	0.927***
	(0.078)	(0.066)	(0.065)	(0.051)	(0.034)	(0.141)	(0.084)	(0.074)	(0.067)
Subject and cohort controls:	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	638	3,100	3,348	2,338	956	638	3,100	3,348	2,338
R-squared	0.016	0.058	0.129	0.200	0.024	0.071	0.118	0.137	0.182
Mean of dep. var. in control	0.807	0.661	0.501	0.264	1.593	-1.517	-0.641	0.158	0.835
P-value: T + T × Girl = 0	0.977	0.019	0.000	0.000	0.147	0.547	0.000	0.000	0.000
Panel B: Newly enrolled sample	**Normalized test score**					**Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	0.153	0.168*	0.425***	0.468***	0.004	0.392	0.307*	0.470***	0.462***
	(0.243)	(0.094)	(0.113)	(0.109)	(0.075)	(0.423)	(0.160)	(0.130)	(0.143)
Girl	0.199	0.093	-0.033	0.122	0.032	0.110	0.054	-0.074	0.136
	(0.221)	(0.060)	(0.076)	(0.081)	(0.045)	(0.318)	(0.099)	(0.119)	(0.090)
T × Girl	-0.485**	-0.028	-0.136	-0.164	0.001	-0.914*	0.032	-0.129	-0.204
	(0.212)	(0.083)	(0.118)	(0.123)	(0.075)	(0.473)	(0.154)	(0.159)	(0.155)
Constant	0.887***	0.683***	0.561***	0.269***	1.667***	-1.108***	-0.526***	0.274**	0.854***
	(0.134)	(0.091)	(0.083)	(0.084)	(0.029)	(0.346)	(0.164)	(0.117)	(0.101)
Subject and cohort controls:	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	43	283	357	315	91	43	283	357	315
R-squared	0.174	0.081	0.234	0.222	0.047	0.412	0.166	0.210	0.182
Mean of dep. var. in control	0.900	0.685	0.346	0.209	1.616	-1.459	-0.742	-0.067	0.766
P-value: T + T × Girl = 0	0.067	0.126	0.003	0.012	0.765	0.148	0.016	0.008	0.123

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1–5 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 6–9 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). Cohort controls consist of indicators for being in the cohort in grade 4 at baseline (grade 3 is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Persistence of Treatment Effects

We next present the effect of the program on test scores at the beginning of the academic year after program exposure. In other words, among those enrolled in 2012, we estimate the effects of the program on the 2013 pre-program test.

Table 6.5 presents the effects of treatment on 2013 pre-program test scores, pooling across test subjects. From Columns 1 and 2 of Panels A and B, we see that there are no average effects in either the Baseline Sample or the Newly Enrolled Sample.[15] We see similar patterns when we look at the probability that a student's test score increased (Columns 3 and 4).

Table 6.5 Treatment effects on 2013 pre-program test scores: Measuring persistence (all subjects)

Panel A: Baseline Sample				
2012 pre-program test score:	**Normalized**		**Probability of improving from pre-program test**	
	(1)	**(2)**	**(3)**	**(4)**
T	0.109	0.208	0.054	0.084
	(0.119)	(0.133)	(0.062)	(0.068)
Girl		0.022		0.015
		(0.066)		(0.034)
T × Girl		-0.188*		-0.055
		(0.098)		(0.049)
Constant	0.000	0.333***	0.550***	0.603***
	(0.077)	(0.095)	(0.039)	(0.046)
	NO	YES	NO	YES
Observations	6,054	6,054	4,325	4,325
R-squared	0.003	0.159	0.003	0.016
Mean of dep. var. in control	0.000	0.000	0.550	0.550
P-value: T + T × Girl = 0		0.874		0.666
Panel B: Newly Enrolled Sample	**Normalized**		**Probability of improving from pre-program test**	
2012 pre-program test score:	**(1)**	**(2)**	**(3)**	**(4)**
T	-0.032	0.209	0.048	0.098
	(0.145)	(0.134)	(0.105)	(0.125)
Girl		0.022		0.084
		(0.067)		(0.100)
T × Girl		-0.190*		-0.113
		(0.099)		(0.125)
Constant	-0.000	0.215**	0.521***	0.511***
	(0.115)	(0.096)	(0.056)	(0.081)
Subject controls:	NO	YES	NO	YES
Observations	1,356	6,054	408	408
R-squared	0.000	0.159	0.002	0.008
Mean of dep. var. in control	0.000	0.000	0.521	0.521
P-value: T + T × Girl = 0		0.874		0.902

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1 and 2 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 3 and 4 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

[15] The absence of impact on learning in the second year holds for all three subjects (results not shown).

Table 6.6 presents these results disaggregated by 2012 pre-program test score. Although the coefficients are relatively large on the normalized test score outcome, the majority of these are not statistically significant at traditional levels. There are small and statistically insignificant coefficients on the probability of improvement outcomes. Consistent with the results for 2013 pre-program tests, the largest coefficients appear in the middle of the pre-test distribution. However, most of these coefficients are not significant.

We also see suggestive evidence of a *negative* impact of treatment on girls' test scores among those with the lowest 2012 pre-program test score (E) in the Baseline Sample (Columns 1 and 6). This negative effect is consistent with the retention results presented previously: the program was effective at retaining more marginal girls in school, but these girls were the least likely to benefit from the program in terms of learning results.

We see no significant effects on the probability of an increase in 2013 pre-program test scores over 2012 pre-program test scores for students with higher pre-program test scores (Columns 6 to 9). In Panel B, we do not find significant effects on average or disaggregated by 2012 pre-program test scores in the Newly Enrolled Sample, but we note small sample sizes in some of these subgroups.

Lastly, we estimate the program's effect on 2013 post-program test scores—in other words, effects experienced by students over two academic years. The overall positive effect observed after one year of program implementation (in 2012, Table 6.3) is no longer significant after two years of program implementation (in 2013, Table 6.7).

Table 6.6 Treatment effects on 2013 pre-program test scores: By 2012 pre-program test scores (all subjects)

Panel A: Baseline Sample	**Dependent variable: Normalized test score**					**Dependent variable: Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	0.130	0.217	0.277*	0.192	0.157	-0.025	0.052	0.130	0.069
	(0.259)	(0.150)	(0.154)	(0.176)	(0.243)	(0.050)	(0.078)	(0.079)	(0.073)
Girl	0.057	0.072	0.004	-0.131	0.281*	0.069*	0.008	-0.008	-0.046
	(0.185)	(0.085)	(0.076)	(0.094)	(0.157)	(0.038)	(0.043)	(0.043)	(0.039)
T × Girl	-0.300	-0.244*	-0.151	0.103	-0.085	-0.138**	-0.053	-0.094	-0.026
	(0.254)	(0.133)	(0.115)	(0.138)	(0.254)	(0.061)	(0.062)	(0.066)	(0.067)
Constant	-0.206	0.059	0.338***	0.612***	0.822***	1.021***	0.807***	0.625***	0.366***
	(0.188)	(0.106)	(0.100)	(0.133)	(0.173)	(0.046)	(0.059)	(0.052)	(0.057)
	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	363	1,497	1,484	981	313	363	1,497	1,484	981
R-squared	0.097	0.154	0.116	0.097	0.034	0.059	0.077	0.099	0.062
Mean of dep, var, in control	-0.601	-0.249	0.087	0.366	0.879	0.955	0.698	0.496	0.263
P-value: T + T × Girl = 0	0.462	0.845	0.327	0.102	0.730	0.004	0.995	0.618	0.556
Panel B: Newly Enrolled Sample	**Dependent variable: Normalized test score**					**Dependent variable: Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	-0.068	0.319	0.407	-0.076	0.085	0.000	-0.012	0.122	-0.031
	(0.196)	(0.267)	(0.289)	(0.310)	(0.239)	(0.000)	(0.134)	(0.156)	(0.132)
Girl	-0.412	0.264	0.125	-0.115	0.152	0.000	-0.108	0.119	0.036
	(0.342)	(0.231)	(0.227)	(0.327)	(0.393)	(0.000)	(0.099)	(0.118)	(0.129)
T × Girl	0.623	-0.490*	-0.506	0.397	-0.223	0.000	0.153	-0.203	0.054
	(0.884)	(0.287)	(0.342)	(0.415)	(0.495)	(0.000)	(0.127)	(0.179)	(0.181)
Constant	1.207**	0.213	0.051	0.504**	1.057***	1.000	1.022***	0.499***	0.362***
	(0.506)	(0.143)	(0.180)	(0.191)	(0.247)	(0.000)	(0.086)	(0.112)	(0.086)
Subject controls:	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	24	112	148	124	21	24	112	148	124
R-squared	0.431	0.251	0.097	0.032	0.075		0.132	0.069	0.038
Mean of dep. var. in control	-0.318	-0.376	0.003	0.315	1.033	1.000	0.727	0.503	0.271
P-value: T + T × Girl = 0	0.529	0.546	0.747	0.399	0.739		0.237	0.643	0.894

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1–5 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 6–9 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Table 6.7 Treatment effect on 2013 post-program test score

Panel A: Baseline Sample				
2012 pre-program test score:	**Normalized**		**Probability of improving from pre-program test**	
	(1)	**(2)**	**(3)**	**(4)**
T	0.155	0.205	0.063	0.090
	(0.117)	(0.126)	(0.059)	(0.062)
Girl		-0.068		0.009
		(0.075)		(0.028)
T × Girl		-0.092		-0.050
		(0.098)		(0.038)
Constant	0.000	0.413***	0.726***	0.785***
	(0.075)	(0.091)	(0.038)	(0.042)
	NO	YES	NO	YES
Observations	5,355	5,355	3,903	3,903
R-squared	0.006	0.210	0.005	0.024
Mean of dep. var. in control	0.000	0.000	0.726	0.726
P-value: T + T × Girl = 0		0.371		0.521
Panel B: Newly Enrolled Sample				
2012 pre-program test score:	**Normalized**		**Probability of improving from pre-program test**	
	(1)	**(2)**	**(3)**	**(4)**
T	0.079	-0.005	0.085	0.060
	(0.147)	(0.172)	(0.076)	(0.104)
Girl		-0.097		0.080
		(0.120)		(0.087)
T × Girl		0.178		0.063
		(0.201)		(0.110)
Constant	0.000	0.450***	0.703***	0.691***
	(0.082)	(0.089)	(0.047)	(0.081)
	NO	YES	NO	YES
Observations	1,212	1,212	350	350
R-squared	0.002	0.199	0.009	0.029
Mean of dep. var. in control	0.000	0.000	0.703	0.703
P-value: T + T × Girl = 0		0.345		0.122

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1 and 2 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 3 and 4 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

We disaggregate the effects by 2012 pre-test scores in Table 6.8. Trends that were visible in the first year continue to be present after the second. Girls previously enrolled with higher initial ability performed significantly higher after the program than their counterparts in the control group, and the largest and most significant effects are found in the middle of the distribution of pre-program scores. More specifically, for the Baseline Sample, the most significant effects are found for boys scoring a C on the 2012 pre-program test and girls scoring a B on that same test.

Table 6.8 Treatment effect on 2013 post-program test score: By 2012 pre-program test score (all subjects)

Panel A: Baseline Sample	**Dependent variable: Normalized test score**					**Dependent variable: Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	0.174	0.195	0.280*	0.150	0.168	-0.016	0.001	0.168**	0.128
	(0.310)	(0.145)	(0.161)	(0.164)	(0.145)	(0.017)	(0.047)	(0.075)	(0.088)
Girl	0.115	-0.081	0.006	-0.125	-0.251	-0.013	-0.043	0.034	-0.055
	(0.235)	(0.088)	(0.100)	(0.077)	(0.292)	(0.016)	(0.029)	(0.048)	(0.039)
T × Girl	-0.387	-0.218	-0.180	0.183	0.279	-0.010	-0.039	-0.113*	0.027
	(0.305)	(0.141)	(0.130)	(0.114)	(0.303)	(0.022)	(0.048)	(0.057)	(0.061)
Constant	-0.185	0.271***	0.420***	0.542***	0.864***	1.024***	0.977***	0.790***	0.570***
	(0.231)	(0.100)	(0.120)	(0.120)	(0.134)	(0.015)	(0.022)	(0.054)	(0.066)
	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	309	1,340	1,356	898	303	309	1,340	1,356	898
R-squared	0.129	0.199	0.158	0.110	0.126	0.023	0.061	0.127	0.102
Mean of dep. var. in control	-0.646	-0.209	0.117	0.343	0.681	0.993	0.879	0.691	0.448
P-value: T + T × Girl = 0	0.307	0.888	0.499	0.038	0.159	0.382	0.512	0.426	0.072
Panel B: Newly Enrolled Sample	**Dependent variable: Normalized test score**					**Dependent variable: Probability of improvement**			
2012 pre-program test score:	**E**	**D**	**C**	**B**	**A**	**E**	**D**	**C**	**B**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**
T	1.155*	0.076	0.577***	0.039	-0.332	0.000	0.002	0.182*	-0.086
	(0.598)	(0.246)	(0.197)	(0.226)	(0.779)	(0.000)	(0.069)	(0.098)	(0.172)
Girl	0.020	-0.042	0.120	0.210	0.197	0.000	-0.095	0.084	0.152
	(0.505)	(0.221)	(0.218)	(0.307)	(0.507)	(0.000)	(0.079)	(0.090)	(0.161)
T × Girl	-1.703**	-0.061	-0.237	0.241	0.590	0.000	0.042	-0.016	0.210
	(0.633)	(0.346)	(0.276)	(0.362)	(0.891)	(0.000)	(0.132)	(0.113)	(0.239)
Constant	-0.052	0.463*	0.135	0.390**	0.622*	1.000	1.046***	0.790***	0.475***
	(0.534)	(0.232)	(0.182)	(0.182)	(0.318)	(0.000)	(0.053)	(0.090)	(0.113)
	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	12	104	125	109	19	12	104	125	109
R-squared	0.785	0.310	0.210	0.126	0.404		0.079	0.190	0.108
Mean of dep. var. in control	-1.261	-0.412	0.022	0.318	0.680	1.000	0.889	0.738	0.452
P-value: T + T × Girl = 0	0.000	0.969	0.209	0.406	0.403		0.714	0.052	0.624

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1–5 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 6–9 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-test score. Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

For the Newly Enrolled Sample, the patterns we see after 2013 are qualitatively similar to those that were observed after 2012. Boys who scored a C on the 2012 pre-program test receive a score that is on average 0.557 standard deviations higher on the 2013 post-program test (Column 3). Further, boys and girls scoring a C on the 2012 pre-program test are both significantly more likely to improve in treatment schools, by 18.2 and 16.6 percentage points, respectively (Column 8).

In sum, we find limited evidence that the strong effects of the program on learning that were observed in the first year did not persist into the second year of implementation. We see this even though some students—those enrolled in third grade in 2012 and fourth grade in 2013—participated in the program over two consecutive years.

School Management

Our last set of results presents the effects of the program on the third prong of the Educate Girls program—school-level SMC outcomes (Table 6.9). Effects each month are presented in Table A.4 in the appendix. Over the course of the seven months in which data were collected, treatment schools held an additional 0.66 committee meetings on average, an increase of 15.6 percent (*p*-value = 0.019). In addition to holding meetings, school committees produced significantly more output, as measured by the number of improvement plans prepared and completed. While the figures are only marginally significant, committees in treatment schools prepared 22.4 percent (*p*-value = 0.086) more improvement plans and completed 24.7 percent (*p*-value = 0.100) more such plans. This finding provides suggestive evidence that an important channel through which learning and retention outcomes may be affected is increased local buy-in.

Table 6.9 Treatment effects on school management committee outcomes (July 2012–January 2013)

Variable	Treatment (1)	Control (2)	Difference (treatment – control) (3)	*P*-value of *t*-test of equality (4)
Number of meetings	4.898	4.239	0.659	0.019
	0.192	0.202	0.278	
Number of improvement plans prepared	8.263	6.752	1.510	0.086
	0.591	0.647	0.875	
Number of improvement plans Completed	5.924	4.752	1.172	0.100
	0.476	0.527	0.708	

Source: Authors' calculations.

Note: Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample for each is 117 treatment and 113 control schools.

7. DISCUSSION

Why Is the Program Impact Weaker for Lower-Performing Students?

In the previous section we showed that the program significantly raised the test scores of all students (girls and boys) after the first year of exposure, except for those at the very bottom of the test score distribution (Table 6.4). Low-achieving students (scoring an E, on an A to E scale) did not benefit from the program.

One component of the Educate Girls program relies on classroom support provided by village volunteers, which effectively lowers the pupil-teacher ratio in low-performing schools. Reducing class size aims to help teachers and volunteers better tailor their instruction to the needs of their students and thus reduce the performance gap. The evidence is mixed as to whether reducing class size is more effective when groups are at mixed levels or when students are tracked. In a context in which teachers lack incentives to focus on high-performing students (most developed countries), positive peer effects make mixed-level groups more effective at improving the scores of initially low-performing students (Epple, Newlon, and Romano 2002). However, if teachers have strong incentives to favor high-performing students, then their behavior is likely to be affected by class composition, and tracking will be necessary to ensure that teachers actually pay attention to low-performing students too (Duflo, Dupas, and Kremer 2011). In many developing countries, teachers' financial rewards are a positive function of test scores, leading teachers to focus their efforts on high-achieving students. Though Duflo, Dupas, and Kremer (2011) argue that teacher payment is likely a convex function of test scores, our results are more in line with a linear function that gives teachers incentives to teach to the top students within each group and explains why, despite tracking, students scoring a D (the second-lowest score) benefited from the program, while students scoring an E (the lowest) did not.

Why Did First-Year Learning Effects Not Last?

The effect of the program did not persist into and after the second year of implementation (Tables 6.5 to 6.8). Studies on long-run effects have shown that the positive effect on test scores of even successful interventions usually fades over time (Banerjee et al. 2007; Andrabi, Das, and Khwaja 2008) due to a tendency to "teach to the test," whereby the program reinforces the specific skills that are required for successful test taking. However, beneficiaries of the program might no longer score higher than others after the intervention stops because skills other than those targeted by the intervention might be required for performing well on subsequent, differently tailored tests. Education interventions are thus more likely to have long-lasting effects when they target core skills (Duflo, Dupas, and Kremer 2011). Accordingly, the fact that the effects of the Educate Girls treatment fade over time corresponds with results of prior studies.

However, the program's positive impacts are not statistically significant and the results suggest smaller effects during the second year, despite the continuation of the Educate Girls intervention. Evidence on other two-year programs has shown even larger effects during the second year than the first—for example, Banerjee and colleagues (2007) found an effect of 0.35 versus 0.18 standard deviations on math test scores and 0.19 versus 0.08 on verbal test scores. One possible explanation for our results is that political economy factors may have changed in the second year and affected volunteers' incentives in the classroom, reducing their overall performance. Volunteers were not paid but offered the possibility of future work. Thus, they might have put in more effort during the first year in order to keep their options open, while in the second year, after a decision was made regarding whether or not they would receive a position with Educate Girls, their effort may have decreased. This explanation is in line with Bold and others' (2013) findings of strong positive impacts of a contract teacher intervention in Kenya when implemented by an NGO (0.18 standard deviations), contrasted with null impact when contract teachers were hired by the Ministry of Education.

8. CONCLUSIONS

This paper presents the results of a three-pronged intervention conducted in rural Rajasthan, India. The program had three primary aims: increasing participation, quality of learning, and gender equity at school among a particularly vulnerable population. We find decidedly mixed evidence of the program's effectiveness at sustainably improving these three outcomes.

First, although enrollment drives specifically targeted marginalized girls, they did not affect either the number or the type of students enrolling in schools. However, the program proved to be successful at retaining students who were already enrolled prior to the start of program implementation, especially girls and low-performing students, across both years of implementation. This temporarily reduced the gender gap in school retention, but these students were then more likely to drop out over the course of the year, limiting the program's overall impact on school attendance.

Second, the large positive effects on learning seen after the first year did not reach marginalized students and tended to fade over time. After the first year of full-fledged program implementation, we see large gains in learning levels, averaging 0.27 standard deviations—more than double the average gains from similar but single-pronged interventions (McEwan 2015) and comparable to the gain in learning resulting from one additional year of schooling. But these first-year benefits were concentrated among students at the high end of the score distribution, and treatment effects after two years of exposure were not significant.

Third, though gender equity in school participation initially improved, in the longer run the program ended up being gender-neutral in terms of enrollment and attendance. Impacts on learning suggest that although girls' learning levels significantly improved after one year of exposure to the program, gender gaps in learning did not diminish. The program's failure to narrow gender gaps may be due to a lack of targeting for the learning component. Once the students were in school, the program's learning intervention lacked a gender-specific component, being available equally to girls and boys. Our results suggest that narrowing gender gaps among this group may require targeting.

In terms of policy recommendations, a program that combines various types of interventions with objectives as diverse as (re-)enrolling marginalized girls in school, increasing learning outcomes for all, and improving school management might dilute service providers' efforts and prevent them from obtaining sustained results.

Methodologically, we provide some evidence that our impact estimates are likely not biased by selection based on observable characteristics. However, survey data collected at home would allow future studies to follow students who were in the original sample and subsequently dropped out of school, thereby ruling out the possibility of bias arising from differences in unobservable characteristics between students in treatment and control schools.

APPENDIX: DATA MATCHING

For the purposes of this project, data were matched from three data sources:

1. Baseline enrollment list, collected in April–May 2011
2. 2012 Creative Learning and Teaching (CLT) test score data
3. 2013 CLT test score data

Datasets 2 and 3 above were matched to 1 by the following hierarchical algorithm (if a match was made at a given step, then those observations were removed from consideration for matches at lower steps):

1. Match if student's name, father's name, school, and grade are consistent.
2. Match if student's name, school, and grade are consistent.
3. Match if student's name and school are consistent.
4. Match if student's name, grade, and village are consistent (to account for within-village transfers).

Due to common names and many missing fathers' names, we encountered a large number of situations in which a given observation from one dataset matched multiple observations from another dataset. To account for such situations, we matched these "multiple matches" in one dataset to all observations in the other dataset in which we found matches, and we created weights to account for the presence of multiple observations in the merged dataset corresponding to single observations in the raw data. For example, if two observations from the 2012 CLT data matched two observations from the baseline enrollment list, we matched these, creating four total matches, and weighted each new "observation" by 0.5. In this way, the weighting scheme ensures that the sum of the weights of any observation from any dataset is 1.0. These weights are used in all results reported in this paper.

Table A.1 Treatment effects on retention to 2013: Conditional on 2012 pre-program test score

Panel A: Baseline sample												
	ALL		**E**		**D**		**C**		**B**		**A**	
2012 pre-program test score (Hindi):	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
T	0.038*	0.046**	0.004	0.068	0.032	0.080***	0.051**	0.033*	0.002	0.019	0.101**	0.032
	(0.019)	(0.018)	(0.078)	(0.047)	(0.028)	(0.025)	(0.023)	(0.018)	(0.032)	(0.023)	(0.039)	(0.029)
Girl		-0.000		-0.024		0.038		-0.007		-0.015		-0.056*
		(0.019)		(0.080)		(0.026)		(0.019)		(0.021)		(0.029)
T × Girl		-0.007		0.022		-0.048		-0.003		0.013		0.067
		(0.023)		(0.089)		(0.032)		(0.024)		(0.025)		(0.040)
Constant	0.426***	0.800***	0.540***	0.772***	0.470***	0.759***	0.412***	0.815***	0.405***	0.824***	0.295***	0.835***
	(0.013)	(0.019)	(0.066)	(0.058)	(0.018)	(0.026)	(0.017)	(0.018)	(0.024)	(0.024)	(0.028)	(0.041)
Subject controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	12,915	12,915	872	872	3,929	3,929	4,233	4,233	2,830	2,830	1,051	1,051
R-squared	0.001	0.673	0.000	0.571	0.001	0.621	0.003	0.700	0.000	0.708	0.011	0.747
Mean of dep. var. in control	0.426	0.426	0.540	0.540	0.470	0.470	0.412	0.412	0.405	0.405	0.295	0.295
P-value: T + T × Girl = 0		0.069		0.296		0.248		0.139		0.209		0.004
Panel B: Newly enrolled sample												
	ALL		**E**		**D**		**C**		**B**		**A**	
2012 pre-program test score (Hindi):	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
T	-0.049	0.035	0.015	0.101	-0.074	0.098*	-0.015	0.057	-0.120	-0.047	0.039	0.062
	(0.050)	(0.039)	(0.174)	(0.187)	(0.077)	(0.058)	(0.055)	(0.047)	(0.075)	(0.068)	(0.111)	(0.099)
Girl		-0.001		0.215		0.043		-0.054		-0.002		-0.035
		(0.053)		(0.174)		(0.076)		(0.070)		(0.069)		(0.101)
T × Girl		-0.040		-0.120		-0.115		-0.006		0.000		-0.108
		(0.063)		(0.246)		(0.087)		(0.083)		(0.091)		(0.154)
Constant	0.429***	0.694***	0.397**	0.496**	0.460***	0.666***	0.425***	0.719***	0.466***	0.730***	0.269***	0.661***
	(0.040)	(0.036)	(0.150)	(0.229)	(0.057)	(0.068)	(0.042)	(0.049)	(0.059)	(0.060)	(0.050)	(0.077)
Subject controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	1,389	1,389	71	71	364	364	468	468	378	378	108	108
R-squared	0.003	0.495	0.000	0.425	0.006	0.488	0.000	0.493	0.015	0.527	0.002	0.525
Mean of dep. var. in control	0.429	0.429	0.397	0.397	0.460	0.460	0.425	0.425	0.466	0.466	0.269	0.269
P-value: T + T × Girl = 0		0.942		0.901		0.810		0.522		0.483		0.718

Source: Authors' calculations.
Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable is an indicator for being enrolled during the 2013 academic year. The sample consists of three observations per student (one for each subject). Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Table A.2 Treatment effects on 2013 pre-program test attendance: Conditional on 2012 pre-program test score

Panel A: Baseline Sample												
2012 pre-program test score:	**ALL**		**E**		**D**		**C**		**B**		**A**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
T	0.050**	0.052**	0.010	-0.036	0.057**	0.068*	0.062**	0.084**	0.002	0.005	0.121***	0.070
	(0.019)	(0.024)	(0.061)	(0.066)	(0.027)	(0.035)	(0.024)	(0.035)	(0.032)	(0.037)	(0.040)	(0.051)
Girl		-0.005		-0.029		0.021		-0.007		-0.011		-0.080*
		(0.023)		(0.086)		(0.035)		(0.028)		(0.025)		(0.047)
T × Girl		-0.003		0.077		-0.022		-0.043		-0.004		0.126
		(0.034)		(0.099)		(0.049)		(0.047)		(0.044)		(0.088)
Constant	0.335***	0.338***	0.426***	0.478***	0.355***	0.351***	0.321***	0.351***	0.340***	0.333***	0.237***	0.291***
	(0.013)	(0.017)	(0.049)	(0.062)	(0.018)	(0.030)	(0.016)	(0.025)	(0.023)	(0.027)	(0.027)	(0.039)
Subject controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	12,915	12,915	872	872	3,929	3,929	4,233	4,233	2,830	2,830	1,051	1,051
R-squared	0.003	0.003	0.000	0.003	0.003	0.004	0.004	0.009	0.000	0.001	0.018	0.026
Mean of dep. var. in control	0.335	0.335	0.426	0.426	0.355	0.355	0.321	0.321	0.340	0.340	0.237	0.237
P-value: T + T × Girl = 0		0.077		0.645		0.223		0.205		0.976		0.006
Panel B: Newly Enrolled Sample												
2012 pre-program test score:	**ALL**		**E**		**D**		**C**		**B**		**A**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
T	-0.070	-0.025	-0.103	0.087	-0.073	0.044	-0.058	-0.080	-0.089	-0.040	-0.070	-0.098
	(0.045)	(0.056)	(0.172)	(0.243)	(0.068)	(0.091)	(0.056)	(0.077)	(0.062)	(0.083)	(0.091)	(0.102)
Girl		-0.007		0.224		0.026		-0.101		-0.025		0.108
		(0.058)		(0.257)		(0.075)		(0.083)		(0.089)		(0.154)
T × Girl		-0.075		-0.443		-0.183*		0.063		-0.081		0.006
		(0.075)		(0.309)		(0.100)		(0.103)		(0.117)		(0.211)
Constant	0.341***	0.344***	0.397**	0.173	0.341***	0.304***	0.347***	0.436***	0.366***	0.358***	0.224***	0.236***
	(0.032)	(0.032)	(0.150)	(0.266)	(0.047)	(0.070)	(0.038)	(0.052)	(0.048)	(0.061)	(0.050)	(0.063)
Subject controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	1,389	1,389	71	71	364	364	468	468	378	378	108	108
R-squared	0.006	0.009	0.012	0.077	0.006	0.026	0.004	0.016	0.009	0.017	0.007	0.051
Mean of dep. var. in control	0.341	0.341	0.397	0.397	0.341	0.341	0.347	0.347	0.366	0.366	0.224	0.224
P-value: T + T × Girl = 0		0.121		0.130		0.095		0.828		0.185		0.613

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable is an indicator for attending the 2012 Creative Learning and Teaching (CLT) post-program test. The sample consists of three observations per student (one for each subject). Subject controls consist of indicators for English and math (Hindi is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Table A.3 Treatment effects on 2012 post-program normalized test scores: By subject

Panel A: Baseline sample	**Dependent variable: Normalized test score**						**Dependent variable: Probability of improvement**					
Test subject	**Hindi**		**English**		**Math**		**Hindi**		**English**		**Math**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
T	0.271***	0.325***	0.247*	0.296*	0.370***	0.429***	0.230***	0.230***	0.201***	0.202***	0.283***	0.272***
	(0.095)	(0.100)	(0.145)	(0.156)	(0.108)	(0.118)	(0.059)	(0.066)	(0.065)	(0.072)	(0.060)	(0.063)
Girl		-0.097*		-0.093*		-0.099*		0.025		-0.003		-0.009
		(0.054)		(0.054)		(0.050)		(0.028)		(0.028)		(0.031)
T × Girl		-0.104		-0.094		-0.114		-0.001		-0.001		0.020
		(0.086)		(0.093)		(0.077)		(0.034)		(0.041)		(0.039)
Constant	-0.000	-0.060	0.000	-0.032	0.000	-0.066	0.577***	0.570***	0.475***	0.481***	0.506***	0.509***
	(0.054)	(0.073)	(0.082)	(0.098)	(0.076)	(0.091)	(0.053)	(0.060)	(0.052)	(0.056)	(0.055)	(0.058)
Cohort controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	4,302	4,302	4,302	4,302	4,302	4,302	2,880	2,880	3,382	3,382	3,162	3,162
R-squared	0.017	0.038	0.013	0.025	0.031	0.052	0.062	0.063	0.042	0.042	0.087	0.088
Mean of dep. var. in control	0.000	0.000	0.000	0.000	0.000	0.000	0.577	0.577	0.475	0.475	0.506	0.506
P-value: T + T × Girl = 0		0.044		0.177		0.005		0.000		0.003		0.000
Panel B: Newly enrolled sample	**Dependent variable: Normalized test score**						**Dependent variable: Probability of improvement**					
Test subject	**Hindi**		**English**		**Math**		**Hindi**		**English**		**Math**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
T	0.302***	0.207	0.283	0.253	0.269*	0.269	0.329***	0.374***	0.226**	0.313***	0.299***	0.428***
	(0.109)	(0.133)	(0.177)	(0.212)	(0.158)	(0.198)	(0.086)	(0.106)	(0.091)	(0.099)	(0.082)	(0.110)
Girl		-0.372***		-0.023		-0.229		0.099		0.134		0.132**
		(0.121)		(0.135)		(0.146)		(0.062)		(0.082)		(0.064)
T × Girl		0.225		-0.011		0.015		-0.094		-0.194*		-0.244**
		(0.193)		(0.182)		(0.204)		(0.092)		(0.104)		(0.115)
Constant	-0.000	0.080	-0.000	-0.113	0.000	0.014	0.483***	0.451***	0.396***	0.328***	0.407***	0.360***
	(0.090)	(0.096)	(0.127)	(0.154)	(0.114)	(0.142)	(0.080)	(0.091)	(0.051)	(0.061)	(0.067)	(0.084)

Table A.3 Continued

Panel B: Newly enrolled sample (continued)	Dependent variable: Normalized test score						Dependent variable: Probability of improvement					
Test subject	**Hindi**		**English**		**Math**		**Hindi**		**English**		**Math**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Cohort controls	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES
Observations	462	462	462	462	462	462	314	314	355	355	329	329
R-squared	0.021	0.049	0.019	0.044	0.018	0.041	0.117	0.123	0.051	0.062	0.090	0.106
Mean of dep. var. in control	0.000	0.000	0.000	0.000	0.000	0.000	0.483	0.483	0.396	0.396	0.407	0.407
P-value: T + T × Girl = 0		0.007		0.187		0.111		0.003		0.283		0.032

Source: Authors' calculations.

Note: Robust standard errors are in parentheses, clustered by village. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in Columns 1–6 is normalized to be mean 0, standard deviation 1. The dependent variable in Columns 7–12 is an indicator for whether the 2012 Creative Learning and Teaching (CLT) post-test score exceeds the 2012 CLT pre-program test score. Cohort controls consist of indicators for being in the cohort in grade 4 at baseline (grade 3 is the omitted category). The Baseline Sample consists of students enrolled at baseline (2011) in grades 3 and 4. The Newly Enrolled Sample consists of newly enrolled students in 2012 and 2013 in corresponding cohorts (grades 4 and 5 for 2012 new enrollees, grade 5 for 2013 new enrollees). All analyses are performed using weights to account for multiple matching.

Table A.4 Treatment effects on school management committee outcomes (July 2012–January 2013): By month

Panel A: School management meetings by month

	Treatment (1)	Control (2)	Difference (treatment – control) (3)	*P*-value of *t*-test of equality (4)
July 2012	0.932	0.841	0.091	0.248
	(0.057)	(0.054)	(0.079)	
August 2012	0.890	0.823	0.067	0.397
	(0.048)	(0.063)	(0.079)	
September 2012	0.737	0.637	0.100	0.234
	(0.061)	(0.058)	(0.084)	
October 2012	0.661	0.602	0.059	0.434
	(0.054)	(0.053)	(0.076)	
November 2012	0.763	0.637	0.126	0.116
	(0.059)	(0.053)	(0.079)	
December 2012	0.788	0.637	0.151	0.033
	(0.049)	(0.050)	(0.071)	
January 2013	0.127	0.062	0.065	0.109
	(0.033)	(0.023)	(0.040)	

Panel B: School improvement plans prepared by month

	Treatment (1)	Control (2)	Difference (treatment – control) (3)	*P*-value of *t*-test of equality (4)
July 2012	1.780	1.531	0.249	0.273
	(0.167)	(0.152)	(0.226)	
August 2012	1.780	1.265	0.514	0.034
	(0.199)	(0.132)	(0.241)	
September 2012	1.127	1.150	-0.023	0.923
	(0.125)	(0.209)	(0.241)	
October 2012	1.237	0.894	0.343	0.098
	(0.149)	(0.143)	(0.207)	
November 2012	1.195	0.796	0.398	0.023
	(0.136)	(0.106)	(0.174)	
December 2012	0.915	1.000	-0.085	0.606
	(0.105)	(0.127)	(0.164)	
January 2013	0.229	0.115	0.114	0.249
	(0.082)	(0.053)	(0.099)	

Table A.4 Continued

Panel C: School improvement plans completed by month	**Treatment** (1)	**Control** (2)	**Difference (treatment – control)** (3)	***P*-value of *t*-test of equality** (4)
July 2012	1.407	1.000	0.407	0.040
	(0.142)	(0.136)	(0.197)	
August 2012	1.263	0.938	0.325	0.090
	(0.149)	(0.117)	(0.191)	
September 2012	0.797	0.805	-0.009	0.967
	(0.101)	(0.187)	(0.210)	
October 2012	0.754	0.549	0.206	0.166
	(0.111)	(0.097)	(0.148)	
November 2012	0.797	0.575	0.221	0.104
	(0.103)	(0.087)	(0.136)	
December 2012	0.720	0.788	-0.067	0.617
	(0.077)	(0.111)	(0.134)	
January 2013	0.186	0.097	0.089	0.301
	(0.070)	(0.049)	(0.086)	

Source: Authors' calculations.
Note: Robust standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample for each month is 117 treatment and 113 control schools.

REFERENCES

Andrabi, T., J. Das, and A. I. Khwaja. 2008. "A Dime a Day: The Possibilities and Limits of Private Schooling in Pakistan." *Comparative Education Review* 52 (3): 329–355.

Angrist, J. D., and V. Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement." *Quarterly Journal of Economics* 114 (2): 533–575.

Baird, S., J. H. Hicks, M. Kremer, and E. Miguel. 2016. "Worms at Work: Long-Run Impact of Child Health Gains." *Quarterly Journal of Economics* 131 (4): 1637–1680.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122 (3): 1235–1264.

Behrman, J. R., S. W. Parker, and P. E. Todd. 2009. "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico." *Economic Development and Cultural Change* 57 (3): 439–477.

Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen. 2013. "Turning a Shove into a Nudge? A 'Labeled Cash Transfer' for Education." *American Economic Journal: Economic Policy* 7 (3): 86–125.

Benveniste, L. A., and P. J. McEwan. 2000. "Constraints to Implementing Educational Innovations: The Case of Multigrade Schools." *International Review of Education* 46 (1–2): 31–48.

Bobonis, G. J., E. Miguel, and C. Puri-Sharma. 2006. "Anemia and School Participation." *Journal of Human Resources* 41 (4): 692–721.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur. 2013. *Scaling Up What Works: Experimental Evidence on External Validity in Kenyan Education*. Working Paper WPS/2013-04. Oxford, UK: Centre for the Study of African Economies.

Borkum, E., F. He, and L. L. Linden. 2012. *The Effects of School Libraries on Language Skills: Evidence from a Randomized Controlled Trial in India*. Working Paper 18183. Cambridge, MA, US: National Bureau of Economic Research.

Bruns, B., A. Mingat, and R. Rakatomalala. 2003. *Achieving Universal Primary Education by 2015: A Chance for Every Child*. Washington, DC: World Bank.

Das, J., S. Dercon, J. Habyarimana, P. Krishnan, K. Muralidharan, and V. Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics* 5 (2): 29–57.

Deininger, K. 2003. "Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda." *Economics of Education Review* 22 (3): 291–305.

Delavallade, C., A. Griffith, and R. Thornton. 2016. "Network Partitioning and Social Exclusion under Different Selection Regimes." Unpublished, International Food Policy Research Institute, Washington, DC.

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

———. 2015. "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools." *Journal of Public Economics* 123: 92–110.

Duflo, E., R. Hanna, and S. P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–1278.

Epple, D., E. Newlon, and R. Romano. 2002. "Ability Tracking, School Competition, and the Distribution of Educational Benefits." *Journal of Public Economics* 83 (1): 1–48.

Filmer, D., and L. Pritchett. 1999. "The Effect of Household Wealth on Educational Attainment: Evidence from 35 Countries." *Population and Development Review* 25 (1): 85–120.

Galiani, S., and P. J. McEwan. 2013. "The Heterogeneous Impact of Conditional Cash Transfers." *Journal of Public Economics* 103: 85–96.

Glewwe, P., M. Kremer, and S. Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–135.

Glewwe, P., M. Kremer, S. Moulin, and E. Zitzewitz. 2004. "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." *Journal of Development Economics* 74 (1): 251–268.

He, F., L. L. Linden, and M. MacLeod. 2008. "How to Teach English in India: Testing the Relative Productivity of Instruction Methods with Pratham English Language Education Program." Unpublished, Columbia University, New York.

Herz, B. 2002. *Universal Basic Education: What Works*. Prepared for the Coalition for Basic Education. Washington, DC: Academy for Educational Development.

Khandker, S. R. 1996. *Education Achievements and School Efficiency in Rural Bangladesh*. Discussion Paper 319. Washington, DC: World Bank.

Khandker, S., M. Pitt, and N. Fuwa. 2003. "Subsidy to Promote Girls' Secondary Education: The Female Stipend Program in Bangladesh." Unpublished, Munich Personal RePEc Archive, Munich, Germany.

Kremer, M., C. Brannen, and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340 (6130): 297–300.

Kremer, M., and A. Holla. 2009. "Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?" *Annual Review of Economics* 1: 513–545.

Kremer, M., and E. Miguel. 2007. "The Illusion of Sustainability." *Quarterly Journal of Economics* 112: 1007–1065.

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

Lavinas, L. 2001. *The Appeal of Minimum Income Programmes in Latin America*. Geneva: International Labour Organization.

Linden, L. L. 2008. *Complement or Substitute? The Effect of Technology on Student Achievement in India*. InfoDev Working Paper 17. Washington, DC: World Bank.

Lloyd, C. B., B. Mensch, and W. Clark. 1998. *The Effects of Primary School Quality on the Educational Participation and Attainment of Kenyan Girls and Boys*. Working Paper 116. New York: Population Council.

Lucas, A. M., P. J. McEwan, M. Ngware, and M. Oketch. 2014. "Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda." *Journal of Policy Analysis and Management* 33: 950–976.

McEwan, P. 2015. "Improving Learning in Primary Schools of Developing Countries: A Meta-analysis of Randomized Experiments." *Review of Educational Research* 85 (3): 1–42.

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

Muralidharan, K., and N. Prakash. 2013. *Cycling to School: Increasing Secondary School Enrollment for Girls in India.* Working Paper 19305. Cambridge, MA, US: National Bureau of Economic Research.

Muralidharan, K., and V. Sundararaman. 2010. "The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India." *Economic Journal, Royal Economic Society* 120 (546): F187–F203.

Petrosino, A., C. Morgan, T. Fronius, E. Tanner-Smith, and R. Boruch. 2012. "Interventions in Developing Nations for Improving Primary and Secondary School Enrollment of Children: A Systematic Review." *Campbell Systematic Reviews* 8: 19.

Pratham Organization. 2011. *Annual Status of Education Report.* Mumbai: Pratham Resource Center.

———. 2012. *Annual Status of Education Report.* Mumbai: Pratham Resource Center.

UNESCO Institute of Statistics. 2016. Net Enrolment Rate, Primary, Both Sexes: India. Accessed January 2017. http://data.worldbank.org/indicator/SE.PRM.NENR?locations=IN.

UNICEF. 2012. *Child Marriage in India: An Analysis of Available Data.* New Delhi.

United Nations. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York.

World Bank. 2001. *Pioneering Support for Girls' Secondary Education: The Bangladesh Female Secondary School Assistance Project*. Washington, DC.

———. 2011. *World Development Report 2012: Gender Equality and Development*. Washington, DC.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go to www.ifpri.org/publications/discussion_papers.
All discussion papers can be downloaded free of charge.

1626. *Limitations of contract farming as a pro-poor strategy: The Case of maize outgrower schemes in upper west Ghana.* Catherine Ragasa, Isabel Lambrecht, and Doreen S. Kufoalor, 2017.

1625. *Cooperation in polygynous households.* Abigail Barr, Marleen Dekker, Wendy Janssens, Bereket Kebede, and Berber Kramer, 2017.

1624. *Farmers' quality assessment of their crops and its impact on commercialization behavior: A field experiment in Ethiopia.* Gashaw Tadesse Abate and Tanguy Bernard, 2017.

1623. *Changing gender roles in agriculture?: Evidence from 20 years of data in Ghana.* Isabel Lambrecht, Monica Schuster, Sarah Asare, and Laura Pelleriaux, 2017.

1622. *Improving the targeting of fertilizer subsidy programs in Africa south of the Sahara: Perspectives from the Ghanaian experience.* Nazaire Houssou, Collins Asante-Addo, and Kwaw S. Andam, 2017.

1621. *El agro Argentino: Un sistema productivo y organizacional eficiente.* Valeria Piñeiro, Miguel Robles, and Pablo Elverdin, 2017.

1620. *Local corruption and support for fuel subsidy reform: Evidence from Indonesia.* Jordan Kyle, 2017.

1619. *Smog in our brains: Gender differences in the impact of exposure to air pollution on cognitive performance in China.* Xi Chen, Xiaobo Zhang, and Xin Zhang, 2017.

1618. *Existing data to measure African trade.* Cristina Mitaritonna and Fousseini Traoré, 2017.

1617. *Misreporting month of birth: Implications for nutrition research.* Anna Folke Larsen, Derek Headey, and William A. Masters, 2017.

1616. *He says, she says: Exploring patterns of spousal agreement in Bangladesh.* Kate Ambler, Cheryl Doss, Caitlin Kieran, and Simone Passarelli, 2017.

1615. *Forced gifts: The burden of being a friend.* Erwin Bulte, Ruixin Wang, and Xiaobo Zhang, 2017.

1614. *Institutional versus noninstitutional credit to agricultural households in India: Evidence on impact from a national farmers' survey.* Anjani Kumar, Ashok K. Mishra, Sunil Saroj, and P. K. Joshi, 2017.

1613. *Cost-effectiveness of community-based gendered advisory services to farmers: Analysis in Mozambique and Tanzania.* Tewodaj Mogues, Valerie Mueller, and Florence Kondylis, 2017.

1612. *The European Union–West Africa Economic Partnership Agreement: Small Impact and new questions.* Antoine Bouët, David Laborde, and Fousseini Traoré, 2017.

1611. *The returns to empowerment in diversified rural households: Evidence from Niger.* Fleur Wouterse, 2017.

1610. *Prospects for the Myanmar rubber sector: An Analysis of the viability of smallholder production in Mon State.* Joanna van Asselt, Kyan Htoo, and Paul A. Dorosh, 2017.

1609. *How do agricultural development projects aim to empower women?: Insights from an analysis of project strategies.* Nancy Johnson, Mysbah Balagamwala, Crossley Pinkstaff, Sophie Theis, Ruth Meinzen-Dick, and Agnes Quisumbing, 2017.

1608. *Stimulating agricultural technology adoption: Lessons from fertilizer use among Ugandan potato farmers.* Lydia Nazziwa-Nviiri, Bjorn Van Campenhout, and David Amwonya, 2017.

1607. *Strengthening and harmonizing food policy systems to achieve food security: A case study and lessons from Ghana.* Suresh Chandra Babu and Sylvia Blom, 2017.

1606. *Trade and economic impacts of destination-based corporate taxes.* Will Martin, 2017.

1605. *Can better targeting improve the effectiveness of Ghana's Fertilizer Subsidy Program? Lessons from Ghana and other countries in Africa south of the Sahara.* Nazaire Houssou, Kwaw Andam, and Collins Asante-Addo, 2017.

1604. *The impact of Ethiopia's Productive Safety Net Programme on the nutritional status of children: 2008–2012.* Guush Berhane, John Hoddinott, and Neha Kumar, 2017.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org