The impact of mentoring and life skills training on secondary school progression and child labor among girls: A randomized controlled trial in Rajasthan

Final Report

May 13, 2019

Abstract
Objectives To measure the impact of a mentoring and life skills training on secondary school progression and child labor among adolescent girls
Design Clustered randomized controlled trial
Setting 119 schools located in the Ajmer District of Rajasthan, India
Participants 2,459 adolescent girls enrolled in grade five in 2016
Intervention An in-school mentoring and life skills curriculum
Main outcome measures Child labor, school progression, life skills
Results Life skills training and mentoring provided in school to adolescent girls by a social mobilizer can improve the expression of life skills by those girls and reduce school dropout rates. Parental assessment of children, child marriage, child labor, and cognitive performance do not change meaningfully.
Conclusions Life skills can be taught in school to adolescent girls, even in areas where girls are relatively disadvantaged. The intervention improves the girls' mutual support as well as helps the girls understand and define goals. These improved life skills help girls stay in school.
Trial registration AEA RCT Registry ID AEARCTR-0001046
Funding Program administered and funded by Room to Read, an education-focused non-governmental organization headquartered in San Francisco with India programs operating out of Delhi. Evaluation funded by a cooperative agreement between American University and the U.S. Department of Labor, IL-26700-14-75-K-25 as well as a grant to Williams College from the Abdul Latif Jameel Poverty Action Lab Post-Primary Education initiative.

Acknowledgments and Disclaimer

We would like to thank a number of people whose efforts, either directly or indirectly, culminated in the completion of this report.

We deeply appreciate all of the support we have received from Room to Read's senior management team and support staff, including Nishant Shrivastava, Padmanav Dutta, Bablu Yadav, Randeep Kaur, Swati Bhardwaj, Nini Mehrotra, Sourav Banerjee, Matthew Jukes, and Ryan Hebert. Their enthusiasm for the study and assistance throughout has been critical to its continued success.

This report would not have been possible without the hard work of Rakesh Pandey, Mohar Dey, and Putul Gupta. We also very much appreciate Arun Chaubey (project associate) and Surya Prakash Upadhyay and Shaitan Singh Gurjar (field managers), who oversaw a large field team of enumerators, as well as the cooperation of local government officials who helped validate the lists of respondents.

Several people were involved in the development of the psycho-social and life skills measures used in this study. We are especially grateful to Eva Roca and Stephanie Jones for their contributions.

The qualitative analysis referenced herein was conducted by Joan DeJaeghere, Aditi Arur, and Devleena Chatterji.

We would also like to thank Lauren Damme and the U.S. Department of Labor's International Labor Affairs Bureau. Their continued support and input at all stages of the project have been invaluable.

This evaluation was conducted in partnership between American University, J-PAL South Asia at the Institute for Financial Management and Research (IFMR), Room to Read, and the United States Department of Labor (USDOL). Each party carried out the following roles in the project:

1. **Room to Read** designed and implemented the mentoring and life-skills program and employed the staff needed to implement the program. For program funding, Room to Read gratefully acknowledges the Bill and Melinda Gates Foundation, Echidna Giving, and Atlassian Foundation International.
2. **J-PAL South Asia at IFMR** is an Indian non-governmental organization that conducts rigorous evaluations of social programs to promote evidence-based policy-making. They were responsible for implementing the evaluation design developed by the American University based team.
3. **American University** assembled the investigators of Eric Edmonds (Dartmouth College), Ben Feigenberg (University of Illinois, Chicago), and Jessica Leight (American University) to conduct the evaluation of the Room to Read Program.
4. **United States Department of Labor (USDOL)** funded this evaluation through the Bureau of International Labor Affairs (ILAB)'s Office of Child Labor, Forced Labor and Human Trafficking. ILAB carries out the department's international responsibilities and funded this evaluation, among others, to build evidence of effective ways to sustainably eliminate child labor around the world.

Funding for this project was provided by the United States Department of Labor under cooperative agreement IL-26700-14-75-K-25. This material does not necessarily reflect the views or policies of the United States Department of Labor, nor does the mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

List of Acronyms

CRC	UN Convention on the Rights of the Child
EFA	Education for All
GEP	Room to Read's Girls' Education Program
IFMR	Institute for Financial Management and Research
SC	Scheduled Caste
SM	Social Mobilizer (woman from the local community providing life skill coaching and mentoring to the girls enrolled in GEP)
ST	Scheduled Tribe
OBC	Other (not scheduled) Backwards Caste
USDOL	United States Department of Labor

EXECUTIVE SUMMARY

While there has been significant progress in promoting schooling and preventing child labor, in much of the world, girls lag behind boys. The barriers that confront girls are multi-dimensional, and ultimately girls need to be empowered to address these barriers. This evaluation considers whether an initiative to build the life skills of adolescent girls in Rajasthan, India can increase the probability that girls progress through school, lower their rates of participation in child labor, and enhance their non-cognitive skills. The project is a randomized controlled trial implemented by American University in partnership with the Abdul Latif Jameel Poverty Action Lab and the non-governmental organization Room to Read.

This evaluation focuses on Room to Read's Girls' Education Program (GEP), a program that has supported more than 95,000 girls in nine countries. GEP is built around a life skills curriculum that begins in grade six and has been developed with attention to the skills and attitudes girls need to unlock their potential, achieve their personal and community goals, and make informed choices about their lives. The curriculum is implemented in school by a "social mobilizer" (SM), typically a young woman from the area, who conducts activities with enrolled girls, including life skills classes, and acts as a mentor and female role model. Relative to the educational timeline, GEP serves girls who have already completed the transition to lower secondary school, and normally continues to serve them until they complete secondary school; however, this evaluation had a two-year time horizon, evaluating the intervention delivered to girls in grades six and seven. As such, the evaluation is not structured to examine longer-term impacts on outcomes such as secondary school completion that are explicit GEP goals or to capture changes in outcomes that take time to accumulate.

This evaluation takes place in the Ajmer district of Rajasthan, India. Room to Read identified schools in Ajmer eligible for the intervention, and J-PAL then selected 119 schools for inclusion in this evaluation. This is the universe of schools in Ajmer that had between 16 and 32 girls enrolled in grade five as of fall 2015, did not have any other NGOs providing life skills curricula to students, and had a classroom in acceptable condition in which a life skills class could take place. The evaluation is based on the analysis of female students who were currently enrolled in grade five in these schools as of January 2016 and who were reached at home and provided consent to be part of the evaluation at baseline (2,459 female students in total).

The evaluation design is based on a stratified, clustered randomization that assigned 60 of the 119 sample schools to receive GEP. Randomization was stratified based on whether schools were above or below median quality, where quality was defined based on a normalized index that included measures of teacher experience, teachers' educational attainment, and classroom and school infrastructure quality. All girls in **treatment** schools received the Room to Read intervention during the school year beginning in June 2016. Girls enrolled in the remaining 59 schools were assigned to the **control** group. All treatment schools continue receiving GEP throughout the evaluation period, and thus all girls enrolled in treatment schools begin receiving GEP in grade six and continue to receive it as long as they stay in school.

A girl is defined as in the treatment group if she was enrolled in grade five in January 2016 in a school that would, barring any household move or other enrollment-related shock, lead to her enrollment in a treatment school for grade six.[1] 94.9 percent of girls in the treatment group enrolled in a treatment school for grade six, and 82.5 percent attended the first life skills training

[1] More specifically, she was enrolled in grade five in a school that also included grades six to eight (and thus would be expected to continue in the same school) and that school was assigned to treatment; or she was enrolled in grade five in a primary school in the same community as a treatment school serving grade six to eight, and thus would be expected to matriculate into the treatment school in the following year.

in their school. By the end of this evaluation in grade seven, 84.7 percent of girls in the treatment group are still engaged with GEP.

This study is designed to answer the following questions:

1. Does life skills education and mentoring delivered by a social mobilizer have an effect on school progression in grades six and seven?

2. Does life skills education and mentoring delivered by a social mobilizer influence the understanding and expression of life skills?

3. Does life skills education and mentoring delivered by a social mobilizer alter child labor among beneficiary girls?

4. Does life skills education and mentoring delivered by a social mobilizer impact the demonstration of cognitive skills and academic achievement?

Figure 1 summarizes this study's findings. In the figure, we present five outcomes related to these hypotheses that illustrate our general findings. For each outcome variable, measured at endline, we present the percentage of the control group for whom that statement is true on the left, followed by the treatment group on the right. 95 percent confidence intervals for the impact of the intervention are also pictured on top of the treatment bar graphs.

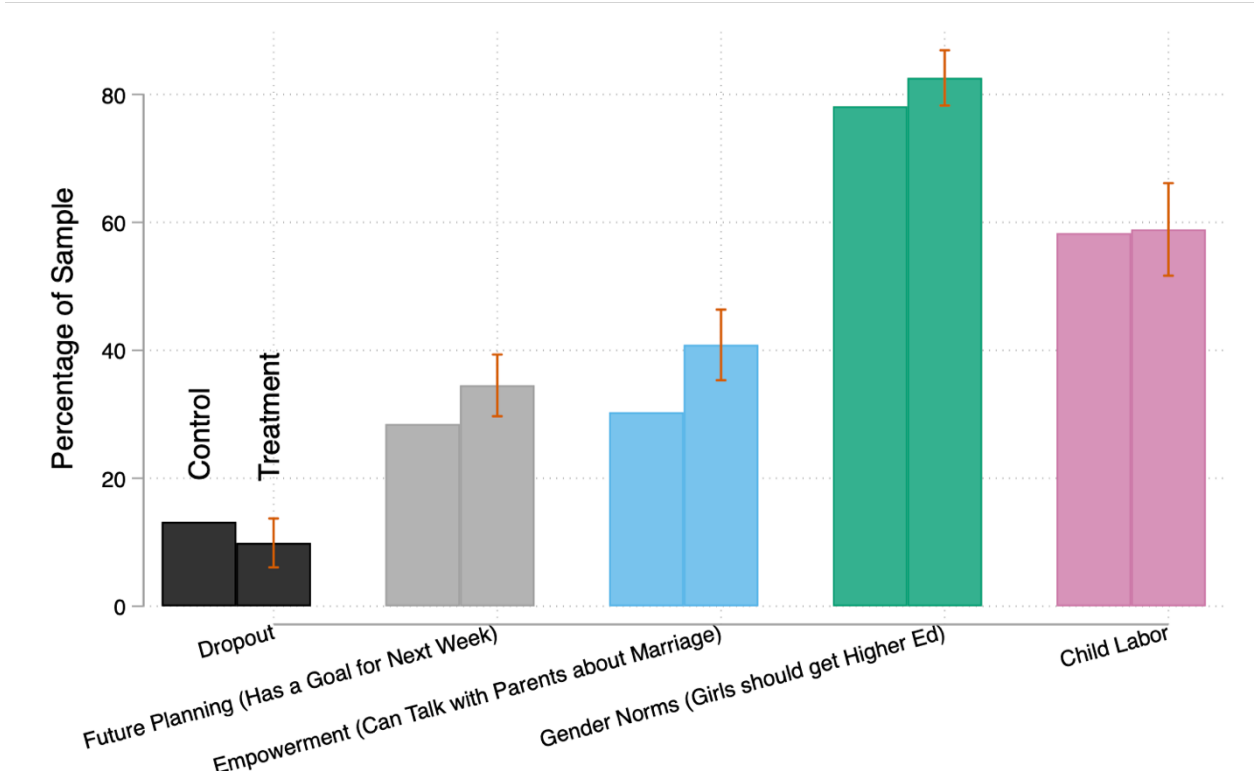

Figure 1: Illustrative Findings

Broadly speaking, we document that GEP improves school progression and the child's expression of life skills. We observe a decline in school dropout with treatment of approximately 25 percent and (not pictured) a 4 percent increase in girls progressing from one grade to the next each year.

The improved expression of life skills stems from a broad range of child responses to questions about their lives and attitudes. The three examples we've pictured in Figure 1 highlight the improvements we document in future planning (shown here as the girl being able to clearly articulate a goal for the next week), empowerment (exemplified in the figure by the girl's ability to talk with her parents about when she will marry), and gender norms (shown in the figure by a girl believing that girls should get higher education even if they marry). These improvements in the child's expression of life skills do not seem to be associated with substantive changes in child marriage or child labor. The final bar in Figure 1 makes clear that the differences observed in child labor are negligible. These improvements in schooling and life skills come without evidence of meaningful improvements in school attendance or cognitive skills (not pictured), measured based on an exam covering language and mathematical knowledge and reasoning. Overall, we answer yes to questions 1 and 2, no to questions 3 and 4.

This study establishes the feasibility of improving the expression of life skills in disadvantaged adolescent girls. While these life skills are beneficial and the overall program improves schooling, they do not appear to measurably influence societal ills like child marriage, child labor, or weak cognitive skills in the time horizon evaluated.

1 INTRODUCTION

1.1 Scientific background and explanation of rationale

Throughout the developing world, there are substantial gender differences in school attendance and child labor. Even where parity in attendance exists, girls continue to be disadvantaged by the curricula, classroom dynamics, teaching methods, and responsibilities outside the classroom. Positive role models can also be scarce in settings with pervasive gender discrimination, and girls often face a variety of hurdles to achieving their potential.

A large literature has analyzed the relationship between school attendance and child labor. Most children who work also attend school (Edmonds and Pavcnik 2005). However, there is evidence of a negative correlation between working and various measures of schooling, including grade advancement, years of education completed, and test scores (Orazem and Gunnarrson 2004, Psacharopoulos 1997 and Ersado 2005). A recent USDOL funded evaluation in Nepal found that incentives for schooling led to an increase in schooling and a decline in carpet weaving (Edmonds and Shrestha 2014, Cooperative Agreement IL-16565-07-75-K), while in another USDOL funded project incentives for attendance and enhanced schooling access in Egypt increased schooling without reducing participation in cotton cultivation (DOLB129K33758). For girls specifically, previous work has suggested that conditional cash transfers can be a successful, but costly, way to increase girls' attendance (Fiszbein and Schady 2009, Dhaliwal et al. 2012), and school-wide provision of free uniforms (Duflo et al. 2012) and bicycles (Muralidharan and Prakash 2013) have been shown to increase the probability of girls completing eighth and ninth grade in Kenya and India respectively.

Financial support can be used to promote schooling and deter child labor, but these tools do not resolve the fundamental disadvantage that girls face in many parts of the world. Edmonds and Shrestha (2014), for example, show that after financial incentives for schooling ended in Nepali carpet factories, girls immediately went back to weaving. Financial support only temporarily influences how children spend their time and does not change the underlying circumstances that generate child labor in the first place. In this study, we are examining whether a more permanent solution is feasible by changing norms rather than using financial incentives to overcome the expression of those norms.

Can empowering girls to develop and express their own agency help them avoid child labor and improve the quality of their lives? This is recognized in many of the key global agreements designed to organize efforts to promote development. The empowerment of girls and women is a target of the Sustainable Development Goal 5. Life skills-based education has the goals of improving the ability of girls to manage day-to-day challenges and enhancing their capacity to make informed, deliberate decisions about their lives. It is a critical input into the empowerment of women, and life skills-based education is recognized as an important component of education in both UN Convention of the Rights of the Child (CRC) and the Education for All (EFA) goals.

However, evidence around the effects of non-financial and life skills interventions targeted to increase schooling and reduce engagement in child labor is not abundant.

This evaluation seeks to answer the question of whether life skills training and mentoring provided by older female role models, denoted "social mobilizers" or SMs, can improve the progress of girls through secondary school, lower their rates of participation in child labor, and enhance their non-cognitive skills. The evaluation is based on a clustered randomized controlled trial implemented by American University in partnership with the Abdul Latif Jameel Poverty Action Lab and the non-governmental organization Room to Read.

Room to Read has been involved in life skills education since 2007 as a part of their Girls' Education Program (GEP). Room to Read's life skills curriculum has evolved over time based on their experiences in the field, involvement of experts in education and psychology, and the inputs of their education and government partners. The curriculum evolves with age starting in grade six and continuing throughout secondary school. This evaluation only captures impacts over the first two years of the program. As such, the evaluation is not structured to examine longer-term impacts on outcomes such as secondary school completion that are explicit GEP goals or to capture changes in outcomes that take time to accumulate. Nonetheless, the evaluation provides useful insights regarding impacts during formative years.

The curriculum aims to develop 10 life skills:

- Self-confidence
- Expressing and managing emotions
- Empathy
- Self-control
- Critical thinking
- Decision-making
- Perseverance
- Communication
- Relationship building
- Creative problem solving

These life skills, or "non-cognitive skills", cover a range of traits that are seen as important determinants of academic achievement, labor market success, and individual well-being. More than 95,000 girls in nine countries have been involved in Room to Read's GEP, and this evaluation of the SM based components of GEP comes at a time where Room to Read anticipates a major expansion of GEP and governments around the world are considering integrating more intensive life skills-based education into schools.

The choice of a clustered randomized control trial to evaluate this program is driven by two considerations. First, because the life skills education is provided to all girls within a school in a given class, it is important that the evaluation be clustered by school. This also mirrors how a life skills education program would be implemented by a government. Second, given Room to Read's inability to finance GEP in all schools, the use of a lottery to allocate the program among equally eligible schools is the simplest and fairest way to allocate Room to Read's limited resources.

1.2 Main Hypotheses

The objective of this project is to evaluate the impact of the life skills education and mentoring parts of the Girls' Education Program implemented by Room to Read. We will examine the impact of this intervention on four primary sets of outcomes, testing four specific hypotheses outlined below. These hypotheses were laid out in the preliminary evaluation design plan shared with the Department of Labor in December 2015. A finalized evaluation design plan is also attached to this report as Appendix 7.2, amended as documented in 7.3. We frame the hypotheses as null hypotheses to match the statistical tests contained in the analysis below.

1.2.1.1 Hypothesis 1: GEP has no effect on school progression or completion of grades six and seven

The primary objective of GEP is to help girls stay in school through secondary school. This objective can be facilitated through the application of the life skills education, the inspiration and support offered by the social-mobilizer mentor, or even the fun and friendships that stem from the program experience. Hence, the impact of the GEP intervention on schooling does not require that life skills are absorbed through the classroom experience. The key outcome measures relevant to this hypothesis include school dropout, progression from one grade to the next, and school attendance.

1.2.1.2 Hypothesis 2: GEP has no effect on life skills

The main components of GEP evaluated here are built around teaching life skills. Hence, the assessment of the program's impact on life skills is informative about whether the program works how it is designed to work and whether students are learning what the life skills curriculum teaches. The key outcome measures relevant to this hypothesis include scaled scores for three objective, task-based measures included in the survey, as well as a number of survey-based measures designed to capture life skills.

- **Objective, task-based measures**. The three task-based measures are: a choice experiment designed to characterize future discounting, a mirror drawing task intended to measure perseverance/grit, and a scavenger hunt designed to measure self-agency as well as perseverance.

- **Survey-based measures**. Key survey-based measures used to evaluate program impacts include: girl's marital status, an index characterizing socio-emotional support**,** an index characterizing freedom of movement, an index characterizing girl's empowerment, an index characterizing girl's self-esteem/self-efficacy, Rosenberg's Self-Esteem Index, an index characterizing girl's future planning, an index characterizing girl's marriage expectations, an index characterizing girl's employment expectations, an index of locus of control, an index of perceived stress, girl's perceptions of gender norms, response to Cantril's ladder question (characterizing subjective life satisfaction), enumerator assessment of girl's behavior during interview, parental perceptions of girl's strengths, parental perceptions of girl's self-efficacy, parental perceptions of girl's freedom of movement, parental perceptions of parent-daughter communication norms, parental perceptions of gender norms, parental attitudes towards girl's schooling, and parental attitudes towards girl's marriage timing.

1.2.1.3 Hypothesis 3: GEP has no effect on child labor

The impact of life skills education and mentoring on child labor is influenced by the intervention's impact on schooling, the capacity of girls and their families to cope with day-to-day difficulties, the ability of girls to plan and follow through on those plans, the desirability of different activities to girls and their families, and the agency girls develop in advocating for themselves. As such, a wide range of impacts on time allocation are possible. The key outcome measures relevant to this hypothesis include participation in hazardous child labor, child labor, economic activity both inside and outside the home, and detailed information on time allocation. Data has been collected to serve as proxies for bonded labor and human trafficking, but we do not expect prevalence rates that would permit further analysis of bonded labor or human trafficking.

1.2.1.4 Hypothesis 4: GEP has no effect on cognitive skills or academic achievement

The same multi-faceted ways that the intervention might influence child labor can also lead to consequences for the demonstration of cognitive skills and academic achievement. The key outcome measures include time spent studying and performance on a cognitive test administered by the research team at endline.

2 METHODS

2.1 Trial Design

2.1.1 Description of trial design

This evaluation is based on a clustered randomized trial with an allocation rate that was intended to be 1:1. In the fall of 2015, 119 schools in the Ajmer district of Rajasthan, India were identified by Room to Read and the research team as appropriate for the intervention. Half received the intervention, leading to a slightly larger allocation ratio because of the odd number of schools.

Prior to randomization, the 119 schools were stratified into two groups based on a school survey conducted by the research team in the fall of 2015. Using information collected in the school survey, the research team created a normalized school quality index, composed of measures of teacher experience, teachers' educational attainment, and classroom and school infrastructure quality. Schools above the median of the index were included in the "high quality" stratum, with the remaining in the "low quality" stratum. School assignment to treatment was conducted separately for the two strata.

Randomization occurred in late 2015, and Room to Read began hiring and training social mobilizers in the spring of 2016. Implementation began in the summer of 2016 with the start of the school year and continues. At the time of design, Room to Read committed to running GEP in treated schools through the school year ending in the spring of 2018, with the goal of continuing GEP in these schools past that date. At the time of writing in the spring of 2019, GEP continues in treated schools and Room to Read plans to continue the program through the end of secondary school.

While randomization is at the school level, the goal of the intervention is to impact girls who attend the life skills education classes and thereby engage with the SM. Hence, the focus of the evaluation is on adolescent girls rather than their schools. Baseline data was collected in the spring of 2016, during which time Room to Read was hiring and training SMs.

The baseline survey involved collecting household information administered through a household survey by the child's caregiver as well as a direct interview of the girl potentially eligible for treatment. Administrative records from schools and Room to Read were also collected throughout the evaluation period.

Following the baseline survey, the sample girls were revisited for tracking surveys in December 2016 and December 2017. The endline survey was conducted between July 2018 and January 2019. Baseline and endline surveys included both a household module and a child module for every girl in the sample, while only girls were surveyed for each of the two shorter tracking surveys.

In addition, qualitative data collection was conducted at baseline, midline, and endline. This involved research activities in six schools served by Room to Read and in the associated communities. Three schools were selected in which school quality is above average, and two schools were selected in which it is below average; a sixth school was selected because it is an all girls' school. The objective of the qualitative data collection is to understand better the channels through which the GEP changes attitudes, perceptions, and decision-making processes for girls, teachers, parents and other stakeholders. Qualitative data was collected by staff members trained in in-depth interview techniques, and collection included the transcription, translation, and coding of the resulting data.

2.1.2 Important changes to methods after trial commencement

Only one significant change to survey methods was implemented following the commencement of the trial. While the original evaluation design called for three tracking surveys, this design was modified to include only two tracking surveys. The request was made to DOL and approved in April 2017 in light of low rates of attrition observed in the first tracking survey; 98% of baseline girls were successfully re-surveyed in the first tracking survey, indicating that two tracking surveys would suffice to minimize attrition over the length of the evaluation.

2.2 Participants

This study takes place in the Ajmer District of Rajasthan. Ajmer was chosen by Room to Read for this study. Room to Read had not operated the current GEP in Ajmer prior to this study, and it was an area where Room to Read wanted to expand because of its location, local capacity, and local school leadership.[2]

2.2.1 Eligibility criteria for schools

The selection of schools eligible for inclusion in this evaluation was undertaken between August and November 2015. A team of enumerators supervised by the research team visited all schools in Ajmer district that included girls enrolled in the relevant grades (six through eight) and collected information about school facilities, staffing, and enrollment. This information was also linked to administrative records about school facilities and enrollment provided by state educational authorities.

The evaluation team and Room to Read then jointly identified criteria that would determine whether or not a school was eligible for inclusion in the evaluation. These criteria included the requirements that the schools enrolled girls in grades six through eight, did not have any other NGOs providing life skills curricula to students, and had a classroom in acceptable condition in which a life skills class could take place. The evaluation team then identified the narrowest possible range of enrollments that would yield a sample of schools enrolling 2500 girls in total; the objective was to have a relatively homogeneous sample of schools in terms of size. This yielded the criteria that the school enrolled between 16 and 32 girls in grade five.

2.2.2 Eligibility criteria for girls

Following the selection of the sample schools, a team of enumerators visited each school between December 2015 and January 2016 to obtain a roster of all girls enrolled in grade five in these schools. All female students who were currently enrolled in grade five in these schools as

[2] A previous, historical version of GEP with a girl-specific design had operated in Ajmer prior to this evaluation.

of January 2016 (2,543 female students in total) were eligible for inclusion in the evaluation. There was no further selection of girls within schools.

2.2.3 Defining the evaluation sample

Every one of the 2,543 girls on the enrollment lists provided by sampled schools was visited at home during the baseline survey; the objective was to conduct a household survey and a child survey for every child in the sample. The baseline survey was conducted before students or their families were informed about the life skills education program.

Ultimately, any girl on the enrollment lists with either a completed household or child survey is considered to be enrolled into the evaluation. Out of the 2,543 female students on the grade five enrollment lists, a total of 2,459 girls from 2,382 households were enrolled into the evaluation sample. Thus, the evaluation is based on 97% of the girls in the sampled school rosters. 2,459 is 98% of our original target sample of 2,500. Not every girl who was part of the evaluation sample was interviewed at baseline. There were 2,353 household surveys conducted at baseline, which provide parent-reported data for 4,237 girls, and 2,399 individual girl surveys conducted at baseline. A flow chart summarizing the sample of girls surveyed and their inclusion in different evaluation phases is included in the Annex as Figure 5.

84 children were on the school enrollment lists but excluded from the evaluation because of failure to complete any component of the baseline survey. 41% were from households that had permanently migrated to a different community prior to the date on which the survey team visited the community– a fact reported by neighbors or other community informants – or simply could not be located. 39% of these 84 girls were excluded because they did not provide consent. The reasons for non-inclusion for the remaining girls varied but included illness or death of the child (4%); parents who were uniformly unavailable during survey hours and thus could not be surveyed or provide consent for the child to be surveyed (3%); and cases in which the child was away from home, particularly during school vacation, and parents declined to participate in her absence (12%).[3]

2.3 Intervention

This evaluation is focused on the impact of social mobilizers on school progression, life skills, child labor, and cognitive achievement. It should be noted that the GEP model that is evaluated in this project is a modified version of GEP as delivered by Room to Read in other contexts; some program components were excluded for the purpose of this evaluation. The modified program includes deployment of social mobilizers who deliver life skills classes and mentoring. The full program additionally includes material support and parent and community engagement. In addition, the full duration of GEP is seven years, serving girls until they graduate from

[3] Percentages do not add up to 100 due to rounding.

secondary school. However, the timeline for this evaluation was two years, following girls during the years corresponding to enrollment in grades six and seven.

40 social mobilizers were employed full-time as a part of this intervention over the lifetime of the intervention, with a maximum of 33 employed at any one time. The typical social mobilizer is responsible for two schools (mean of 1.95). GEP aims to have 50 girls per SM. SMs are 33 years of age on average. All the social mobilizers had completed both secondary and post-secondary education, and all were from Ajmer district; within the district, 58% were from urban areas. Prior to beginning in the classroom, SMs receive 14 days of training with an additional eight days of training at the start of each subsequent school year. Every eight SMs have a program assistant for supervision and support, who in turn is supervised and supported by a Senior Program Associate and Senior Program Officer.

The social mobilizers engage in two primary activities within the scope of Room to Read's GEP. During the school day, SMs provide life skills education to girls enrolled in the program. SMs also provide mentoring to their students. The goal of both these activities is to help girls develop life skills to negotiate schooling, prioritize education, develop agency, and prepare for life. Figure 2 provides the logic model of the intervention.

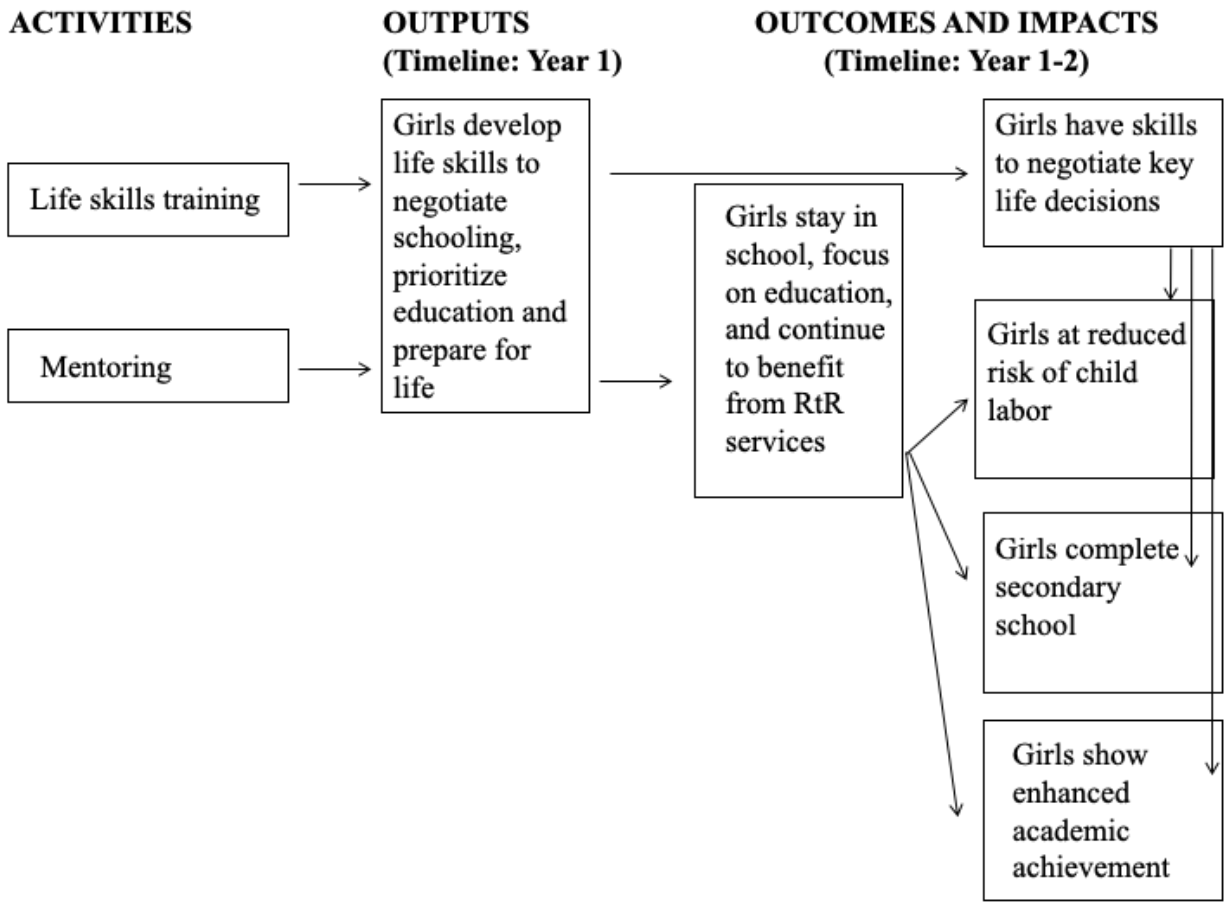

Figure 2: Logic Model

Life skills training takes place within the school during the school day. Room to Read developed the life skills curriculum. While there is some localization, it is similar across all the countries where Room to Read runs GEP. The curriculum is grade-based and emphasizes self-confidence, expressing and managing emotions, empathy, self-control, critical thinking, decision-making, perseverance, communication, relationship building, and creative problem solving. The intervention also focuses on applying these skills to simulations involving time management, education, physical protection and rights, health, and community involvement. It evolves as girls age (starting in grade six and ending at the end of secondary school) and regularly revisits topics, adapting to stay age appropriate and relevant.

The life skills training was completed for grades six and seven in our evaluation. Life skills classes are held once every other week. Every treatment school has 16 life skills classes conducted in both grades.

Because life skills classes are held during the school day, ordinarily a student needs to attend school to attend the life skills class. Students who miss a class are not excluded from subsequent

classes. Figures 3 and 4 contain a histogram of the fraction of life skills classes attended by subjects in grade six and seven, respectively.

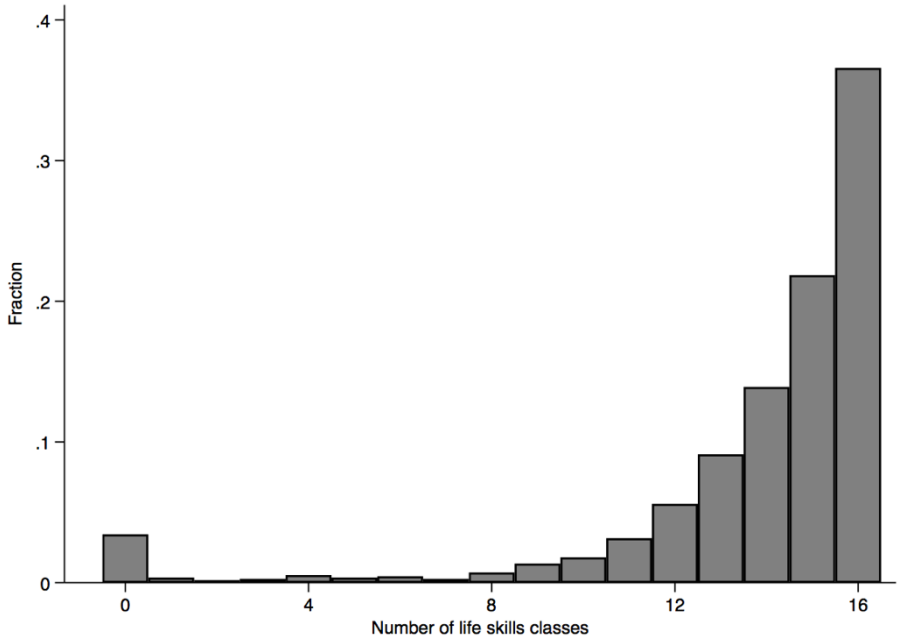

Figure 3: Number of Life Skills Classes Attended by Treatment Group Subjects in Grade 6

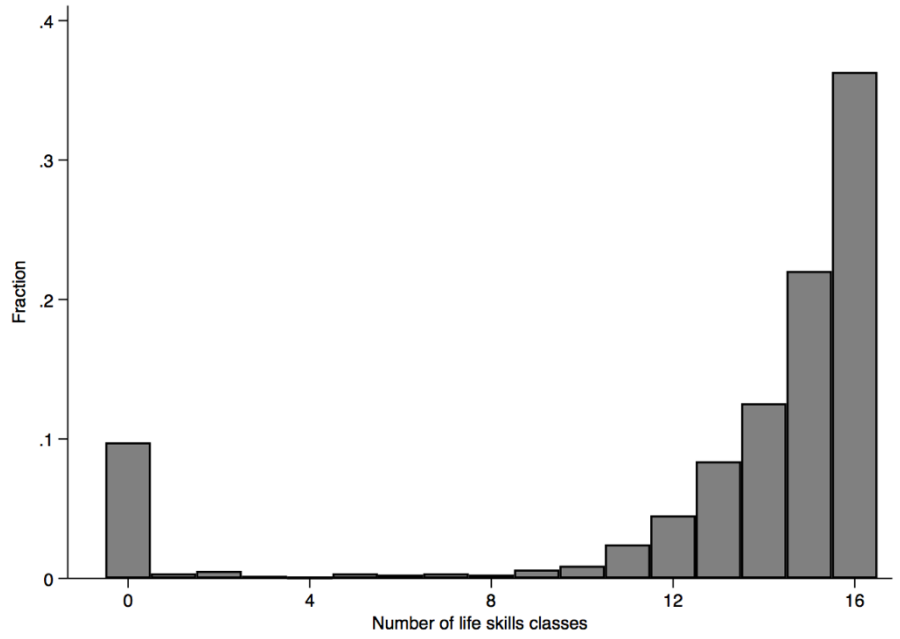

Figure 4: Number of Life Skills Classes Attended by Treatment Group Subjects in Grade 7

In addition to life skills sessions, the intervention entails monthly group mentoring sessions conducted by the social mobilizer. In preparation for these mentoring sessions, SMs receive training in recognizing risks in girls and orientation on additional support services that might be suitable for helping girls triage risks. The sessions are aimed at helping the girl troubleshoot difficulties in her life, develop long-term goals in schooling and career, enhance her agency, and boost her self-confidence in life skills areas needing additional work.

Of the 40 SMs involved in this evaluation, seven left at some point over the two years and were replaced within three months. One was released for poor performance, and others left for personal reasons such as marriage or migration. Each SM was observed quarterly to assess the quality of her life skills session and to provide her with support to improve session delivery.

In qualitative interviews conducted with a subset of girls participating in GEP and their parents, respondents often highlighted the effort put forth by social mobilizers to form strong and supportive relationships with participating girls. These interviews also highlighted the degree to which the quality of assigned social mobilizers impacted the perceived efficacy of the intervention.

2.4 Outcomes

2.4.1 Completely defined pre-specified primary and secondary outcome measures

The pre-specified evaluation design plan, shared with the Department of Labor in December 2015, outlines analyses related to treatment effects on four primary sets of outcomes, corresponding to the four numbered hypotheses specified in section 1-B above.

Relevant outcome variables were all defined in the evaluation design plan attached to this report as Appendix 7.2. The majority of outcome variables are measured in both the baseline and endline surveys; the question numbers referred to below correspond to the relevant question in the baseline survey.

No secondary outcomes were specified. We also did not pre-specify any analyses employing data collected from two tracking surveys undertaken. These tracking surveys were designed primarily to increase the frequency of contact with girls in the study sample in order to minimize endline attrition.

2.4.1.1 Hypothesis 1: GEP has no effect on school progression or completion of grades six and seven

For the first hypothesis, four outcome variables of interest are defined.

- School dropout: This will be measured based on an indicator variable equal to “1” if either household survey Q214 response is “No” for the girl, or if child survey Q203 response is “No.”

- Grade progression: This will be measured based on an indicator variable equal to "1" if girl is enrolled at or above grade level seven based on response to child survey Q204. Variable will be coded as "0" for girls not enrolled in school or girls who have not advanced to grade seven.
- School attendance: This will be measured based on responses to child surveys Q205 and Q206. Two attendance measures will be constructed.
    - The first divides Q205 by Q206 to give the fraction of the days that schools was open in the last week where the child attended.
    - The second is an indicator that is 1 if the child reports attending school in the last week (Q205>0).
    - Both variables will be coded as missing if school was not open in the last week or the child dropped out.

2.4.1.2 Hypothesis 2: GEP has no effect on life skills

For the second hypothesis, four objective measures and 21 survey-based variables of interest are defined.

The objective measures include: a choice experiment designed to characterize future discounting, a mirror drawing task intended to measure perseverance/grit, and a scavenger hunt designed to measure self-agency as well as perseverance. These four measures are constructed as follows.

- Future discounting: Responses to child survey Q510a will characterize whether a girl reports preferring a given payment at present or a larger payment in the future.
- Mirror drawing: Two variables will be constructed to measure task performance.
    - The first variable will be equal to the number of mirror drawings successfully completed (up to four).
    - The second variable will measure the total number of seconds spent on all mirror drawing tasks. It will be missing if no mirror drawing tasks were attempted.
- Scavenger hunt: An index will be created from two normalized measures. The first measure will count the total number of items successfully collected. The second measure will count the number of items that the girl attempted to collect (whether successful or not).

The survey-based measures include: girl's marital status, an index characterizing socio-emotional support**,** an index characterizing freedom of movement, an index characterizing girl's empowerment, an index characterizing girl's self-esteem/self-efficacy, an index characterizing girl's future planning, an index characterizing girl's marriage expectations, an index characterizing girl's employment expectations, girl's perceptions of gender norms, response to Cantril's ladder, enumerator assessment, parental perceptions of girl's strengths, parental perceptions of girl's self-efficacy, parental perceptions of girl's freedom of movement, parent-daughter communication, parental perceptions of gender norms, parental attitudes towards girl's

schooling, parental attitudes towards girl's marriage timing, an index of locus of control, an index of perceived stress, and Rosenberg's Self-Esteem Index.[4] These 21 measures are constructed as follows.

- Girl's marital status: An indicator variable for whether girl is married or committed (engaged). This variable will be equal to "1" if household survey Q206 takes on values "2"-"4" for the girl.
- Socio-emotional support: A normalized index will be generated based on responses to child survey Q501a-Q502c.
- Freedom of movement: A normalized index will be generated based on responses to child survey Q503c1-Q503c7. Responses to each question will be coded as "1" if girl reports that she is allowed to go alone or accompanied and coded as "0" otherwise.
- Empowerment: A normalized index will be generated based on responses to child survey Q504a-Q505h. Q504a-Q504k responses will take on one of three possible values, while Q505a-Q505h will take on one of two possible values. Each variable will be coded so that higher values correspond to greater levels of empowerment/autonomy.
- Self-esteem/self-efficacy: A normalized index will be generated based on responses to child survey Q507a-Q507d, Q508c-Q508g and Q516a-Q516b; responses will take on one of two possible values. "Don't Know" responses to Q507a-Q507d will be coded as missing. "Have not thought about this" responses to Q508c-Q508g will be coded as equivalent to "No".
- Future plans: A normalized index will be generated based on responses to child survey Q508a-Q508b, Q511a, Q512c, Q512e-Q512f. For Q511a and Q512c, indicator variables will be created based on whether any goal is provided (i.e. responses #1-#6 given). For Q512e-Q512f, responses will take on one of five values, with the highest value corresponding to "Very true for me" and the lowest value corresponding to "Have never thought about goals."
- Marriage expectations: A normalized index will be generated based on responses to child survey Q513c-Q513f along with the difference between Q513a and Q513b. "Never" responses to Q513a will be coded as missing. For girls already married, the index will be set equal to the lowest value calculated for a non-married respondent.
- Employment expectations: A normalized index will be generated based on responses to child survey Q514a, Q514c-Q514d. "Other" responses to Q514c, Q514d will be coded based on the closest corresponding numerical response value.
- Gender norms: A normalized index will be generated based on responses to child survey Q515a-Q515f.

[4] Cantril's ladder is a measure of life satisfaction (Cantril 1965). The Rotter locus of control index measures the extent to which respondents believe they have control over life events, as opposed to life events being controlled by external forces (Rotter 1966). The index of perceived stress is a widely used psychological instrument for measuring the perception of stress (Cohen 1983). Rosenberg's index is a commonly-used index characterizing global self-worth (Rosenberg 1965).

- Cantril's ladder: Responses will be coded from 1 to 10.
- Enumerator assessment: A normalized index will be generated based on responses to child survey Q702-Q705.
- Parental perceptions of girl's strengths: A normalized index will be generated based on responses to household survey Q416a-Q416u.
- Parental perceptions of girl's self-efficacy: A normalized index will be generated based on responses to household survey Q417a-Q417e. Responses will be coded from "1" to "5". "Don't Know" responses will be coded as missing.
- Parental perceptions of girl's freedom of movement: A normalized index will be generated based on responses to household survey Q418a-418h. Responses to each question will be coded as "1" if respondent reports that girl is allowed to go alone or accompanied and coded as "0" otherwise.
- Parent-daughter communication: A normalized index will be generated based on responses to household survey Q419a-420h. Responses to Q419 take on one of two values, while responses to Q420 take on one of three values.
- Gender norms: A normalized index will be generated based on responses to household survey Q424a-424i.
- Parental attitudes towards girl's schooling: A normalized index will be generated based on responses to household survey Q425-Q427 and Q440b. "Other" values will be coded corresponding to the closest numerical value and "Don't Know" responses for Q440b will be coded as missing.
- Parental attitudes towards girl's marriage timing: A normalized index will be generated based on (1) difference between household survey Q434 and Q432 responses, (2) difference between household survey Q435 and Q432 responses, (3) Q437, and (4) Q440d-Q440f. For Q440d-Q440f, "Don't Know" and "I do not expect her to do this" responses will be coded as missing.
- Rotter locus of control: A normalized index will be generated based on responses to Q518. The scoring follows established scoring methods for this module, assigning one point for each response that the respondent selects that is consistent with an external locus of control. The total score is the sum of scores on all items.
- Perceived stress: A normalized index will be generated based on responses to Q519. The scoring follows established scoring methods for this module, entailing a Likert scale of 0 to 3 for each item; the total score is the sum of scores on all items, ranging from 0 to 12.
- Rosenberg self-esteem index: A normalized index will be generated based on responses to Q520. The scoring follows established scoring methods for this module, entailing a Likert scale of 0 to 3 for each item; the total score is the sum of the scores on all items, ranging from 0 to 30.

For all life skills indices described above, we will first take the difference between each component survey response value and the mean within the control group and then divide by the

control group standard deviation. We will then average over all index components, ensuring that values for each component are constructed so that the index interpretation is consistent (i.e. higher values of empowerment index components all correspond to higher levels of empowerment). Note that at each analysis phase, the control group mean and standard deviation will be calculated from the concurrent data (i.e. control group mean and standard deviation will be calculated in the endline data when outcomes from the endline data are being analyzed).

2.4.1.3 Hypothesis 3: GEP has no effect on child labor

For the third hypothesis, ten outcomes of interest are defined.

- Working child
  - Children who have worked in family-based work in the last 12 months.
  - Children engaged in child labor.
- Working for pay
  Children who report being paid for their work in the last 12 months regardless of the location of the work. Pay includes cash, shelter, food, clothing, or medical support. Reports of pay in last 7 days will be combined with questions governing the last 12 months.
- Working outside of family-based activity
  Children who report working outside of a family-based activity in the last 12 months (including responses with last 7 days recall). An activity is no longer considered family based if employees are present.
- Child labor
  - Children working outside of family-based work in the last 12 months.
  - Children economically active for 20 or more hours in the last week.
  - Children participating in unpaid household services for 20 or more hours in the last week.
  - Children engaged in hazardous child labor.
  - Children engaged in other worst forms of child labor.
- Hazardous child labor
  - Children in an occupation listed in Appendix One.
  - Children participating in any of the processes listed in Appendix Two.
  - Children working in an economic activity in the last 12 months that is hazardous by its nature or circumstance (based on self-reported job characteristics in questions 401-415).
- Other worst forms of child labor
  - Children (potentially) in debt bondage (reports working for a debt and being unable to leave).
  - (Potentially) Trafficked Children (reports coming from a different community with a trip organized by a non-parent or with other working children and being unable to leave).
  - Child (potentially) in forced labor (reports being forced to work and unable to leave).
- Hours working: A continuous variable measuring hours spent in economic activity, including collection activities and commuting time, during a typical day in the last week based on child survey Q311 time allocation responses.

- Hours in unpaid household services: A continuous variable measuring hours spent in unpaid household services during a typical day in the last week based on child survey Q311 time allocation responses.
- Active: A continuous variable measuring hours spent in unpaid household services and working during a typical day in the last week based on child survey Q311 time allocation responses.
- Hours active outside of the home: A continuous variable measuring hours spent in unpaid household services and working outside of the child's home during a typical day in the last week based on child survey Q311 time allocation responses.

2.4.1.4 Hypothesis 4: GEP has no effect on cognitive skills or academic achievement

For the fourth hypothesis, five outcomes of interest are defined.

- Time spent studying: A continuous variable measuring minutes spent on school work outside of school during a typical day in the last week based on child survey Q311 time allocation responses.
- Hours spent on school: A continuous variable measuring hours spent on school including time spent studying, time in school, and time traveling to and from school during a typical day in the last week based on child survey Q311 time allocation responses.
- Cognitive test scores: The student's score on the ASER tests in mathematics, Hindi and English included in Section 6 in the survey.[5] The test scores correspond to three separate variables for each assessment module.

2.4.2 Any changes to trial outcomes after the trial /commenced, with reasons

There were three primary changes made to the pre-specified trial outcomes after the evaluation commenced; these were reported to the funder and added to the evaluation's AEA registration entry as of August 2018.

The first change was the addition of three non-cognitive indices: the Rotter locus of control module (now Q518), a module focusing on perceived stress (now Q519), and the Rosenberg self-esteem module (now Q520). All three modules have been developed in other contexts and used extensively. The objective of including these additional indices was to ensure that our evaluation also reported on non-cognitive measures that are employed in other relevant evaluations, thus enhancing the external validity of our results. For example, Heckman, Stixrud and Urzua (2006) employs a noncognitive index based on both the Rotter locus of control and Rosenberg self-esteem scales and identifies a strong link between noncognitive skills and schooling as well as wage outcomes. Coleman and DeLeire (2003) also uses a variant of the Rotter locus of control instrument to identify a similar relationship between a belief in the importance of "internal" factors in affecting life events and human capital acquisition. In the Indian context, Joshi and

[5] ASER (Annual Status of Education Report) tools are provided by Pratham, an educational non-governmental organization in India, and can be found on-line at www.asercentre.org.

Srivstava (2009) demonstrates lower levels of self-esteem among girls based on the Rosenberg survey instrument. The index of perceived stress that we employ has been used previously to document various correlates of perceived stress, including female gender and low subjective socioeconomic status (see, for instance, Hamad et al., 2008).

The second change was the specification of the cognitive test to be employed. The original evaluation design plan simply noted that we would be analysing scores on a cognitive test; we subsequently opted to use the ASER test in the endline survey following extensive exploration of appropriate instruments as well as piloting.

The third change related to the measurement of future discounting. After initial piloting, we came to the conclusion that sample girls were not able to consistently respond to a series of included questions related to preferences for current versus future payoffs that varied the relative payoff amounts. In response, we collapsed the original (six) questions related to future discounting to a single question and included this question in both baseline and endline surveys.

2.5 Sample Size

2.5.1 How sample size was determined

The evaluation sample includes 119 schools; this is all schools in Ajmer meeting Room to Read's eligibility criteria that had between 16 and 32 girls enrolled in grade five as of fall 2015. The sample size was jointly determined by the research team and Room to Read in order to maximize statistical power within the constraints of the available budget for program expansion.

Power calculations for the evaluation were constructed in the study design phase using data on school achievement and child labor in Rajasthan; these power calculations were subsequently updated using baseline data and presented in the baseline report.[6] The enrollment, child labor, and marriage rates for girls 12-14 (the estimated age of the sampled cohort at follow-up) were constructed based on the data from older siblings/cousins of the girls in the study sample (i.e., other girl residents in the same household) given that the objective was to predict outcome variables when the girls in the target population reached the age of 12-14 (at the time of the endline survey).

Given a measured enrollment rate of girls between ages 12 and 14 of 90 percent, the planned evaluation size (60 treatment schools and 59 control schools) would allow us to detect an increase in this enrollment rate to 95 percent, a proportional increase of 6 percent. Given a measured baseline participation rate of girls in work for pay of 82 percent, the planned evaluation would allow us to detect a decline in this participation rate to 73 percent, a proportional decrease of 11 percent. Given a measured child labor rate of girls of 89 percent, the planned evaluation would allow us to detect a decline in this child labor rate to 81 percent, a

[6] The power calculations were conducted in Stata utilizing the command clustersampsi.

proportional decrease of 9 percent. Given that 28 percent of surveyed secondary school-aged girls are married in our households (as are 17 percent of our subjects), the planned evaluation would allow us to detect a decline in this marriage rate to 18 percent, a proportional decrease of 36 percent.

This trial did not employ any interim analysis or stopping guidelines.

2.6 Randomization

2.6.1 Sequence generation

Randomization was conducted by the research team using Stata. More specifically, a stratified randomization was conducted assigning 60 of the 119 sample schools to the treatment group. Randomization was stratified based on whether schools were above or below median quality, where quality was defined based on a normalized index that included measures of teacher experience, teachers' educational attainment, and classroom and school infrastructure quality.

Following the initiation of data collection, it was discovered that three of the schools selected to be in the sample in fact did not enroll girls past grade five; for the upper-level grades, these were single-sex schools including only boys. During the sample selection process, these schools were incorrectly designated as including higher-grade girls as well. These three schools (two treatment and one control school) were dropped, and an additional three schools were selected to replace them.[7] These replacement schools constituted an additional third strata.

2.6.2 Allocation concealment mechanism

Since school-level treatment assignment was determined prior to any survey activities (and the collection of consent from survey participants), no steps were taken to conceal the results of the randomization.

2.6.3 Implementation

As noted above, randomization was conducted by the research team, and the list of treatment schools was then communicated to Room to Read. Room to Read was then responsible for enrolling students into the intervention in intervention schools.

The evaluation team enrolled individual girls and households into the evaluation sample at baseline using a detailed process of consent administered for both household and child surveys. Enumerators were trained to explain the purpose of the study, the benefits of participating, the study's duration, and the frequency of the proposed interviews. Interviews were conducted only

[7] The replacement process for these schools entailed identifying 12 schools that met the eligibility criteria if the enrollment window was slightly lowered to 15. Three schools were randomly chosen to join the sample among the 12, and of these, 2 were randomly assigned to the treatment group.

after respondents consented to participate and all questions regarding the study were addressed. Separate consents, both verbal and written, were obtained from the members who participated in the household survey. For the child survey, parental consent from the primary caregiver was first obtained before interviewing the child. In case the primary caregiver of the child was not available, consent was obtained from the most senior member of the household. Informed verbal consent was obtained from all children participating in the study. The consent process was then repeated for each subsequent survey.

2.7 Blinding

Given the nature of the intervention, participants were not blind to their treatment assignment. Parents, community members, teachers and other stakeholders were aware whether or not girls in their households or communities were offered Room to Read programming.

Within the research team, the principal investigators and the research assistant and research manager leading the field team were similarly not blinded to treatment assignment. This is primarily because they also had the responsibility of monitoring the fidelity of intervention implementation and adherence by the partner to the randomized design. However, enumerators and supervisors conducting and overseeing surveys were blinded to treatment assignment of households.

2.8 Statistical Methods

2.8.1 Statistical methods used to compare groups for primary and secondary outcomes

To identify the impacts of the intervention, we employ three specifications. First, we estimate an ordinary least squares (OLS) regression of each outcome of interest on an indicator variable for treatment assignment and indicator variables for randomization strata. No additional baseline control variables are included. The equation of interest can thus be written as follows for each child outcome, denoted Y_{ist} for child i in school s measured at time t. T_s denotes the dummy for treatment assignment, and μ_s denotes the randomization strata for school s. This equation will be denoted specification (1).

$$Y_{ist} = \beta_1 T_s + \mu_s + \varepsilon_{ist}$$

Second, we estimate a specification including baseline control variables in which the outcomes of interest are regressed on an indicator variable for treatment assignment, indicator variables for randomization strata, a vector of age dummies, a vector of dummies capturing the most important type of employment in the household at baseline, and a control variable that measures the lagged (baseline) value of the relevant outcome. The equation of interest can thus be written as follows; $Y_{is,t-1}$ denotes the baseline value of the outcome, γ_i denotes a vector of age dummies, and λ_i denotes a vector of indicator variables for the most important type of employment in the household at baseline. This equation will be denoted specification (2).

$$Y_{ist} = \beta_1 T_s + \beta_2 Y_{is,t-1} + \mu_s + \gamma_i + \lambda_i + \varepsilon_{ist}$$

The inclusion of the lagged control measure will serve to improve precision of estimated treatment effects as will the age and employment type controls. For the family of outcomes corresponding to school progression and completion (Hypothesis 1), age at enrollment and maternal education are additionally included as control variables.

If the baseline control variable is missing because either the household or child survey was not conducted for a particular girl at baseline, the missing value is coded as zero. Additional dummy variables equal to one for observations with missing values are included for each baseline covariate. This methodology for addressing missing baseline values was pre-specified in the evaluation design plan addendum.

Third, we estimate a specification that adds to specification (2) controls for baseline variables where imbalance was reported between the treatment and control households in Tables 3 and 4. These baseline variables are denoted $\chi_{is,t-1}$. This equation will be denoted specification (3).

$$Y_{ist} = \beta_1 T_s + \beta_2 Y_{is,t-1} + \beta_3 \chi_{is,t-1} + \mu_s + \gamma_i + \lambda_i + \varepsilon_{ist}$$

A small number of outcome measures will not have baseline data available. As previously noted, three new non-cognitive indices were added at endline (the Rotter locus of control index, the perceived stress index, and the Rosenberg self-esteem index). In specifications examining these outcomes, we will control for lagged values of overall life skills indices since lagged outcome values will be unavailable. In addition, a cognitive test (the ASER test developed by Pratham) was added at endline. In specifications examining this outcome, we will control for baseline school dropout status, attendance, grade progression, time spent studying, hours spent on school and reported grades in grade five as reported in administrative data. This methodology was also pre-specified in the evaluation design plan addendum.

In all specifications, standard errors will be clustered at the school level. Our sample includes 119 clusters. There are a large volume of hypotheses tested regarding life skills. With so many hypothesis tests, there will be false discoveries (type 1 error). For all life skill measures, we will also compute false discover rate adjusted q-values across all life skill outcomes using the same specification (Benjamini and Hochberg (1995)).

2.8.2 Additional analyses

For all outcomes of interest, we pre-specified at baseline that we would conduct heterogeneity analysis based on child grade-for-age at baseline, mother's education, school quality, and baseline cognitive test performance. However, given that baseline cognitive test scores were ultimately not collected, heterogeneity analysis based on cognitive test performance will not be conducted.

The heterogeneity analysis will be implemented as follows: in addition to including an indicator for treatment assignment, the first such specification will include the interaction of the indicator for treatment assignment with a discrete variable measuring whether the child's school was in the high school quality stratum (above median quality based on the index described in 2.1.1), the second specification will interact treatment assignment with child's age, and the third specification will interact treatment with a discrete variable indicating whether the mother completed primary. Finally, we will examine heterogeneous impacts based on whether the household has experienced the following types of household shocks: (1) economic shocks, (2) household illness or death, and (3) other shocks including crime and land/family disputes (occurrence of shocks is measured in household survey Q111 and Q113). All heterogeneity analysis specifications will include a variable that controls for the source of heterogeneous treatment effects (i.e., a control for child grade-for-age).

2.8.3 Qualitative Analysis

Extensive qualitative work is an important part of this study. A research team led by Joan DeJaeghere of the University of Minnesota targeted 60 girls in total from six different schools in Rajasthan three times: before the start of the program when girls were in grade five, at midline after grade six, and at endline after grade seven. The interviews were semi-structured, open ended interviews and involved the girls, their parents, teachers, and social mobilizers. There was some attrition as six girls were lost over the three year period of qualitative interviews. We incorporate these qualitative findings in our discussion section 4.

3 Results

3.1 Participant flow

Figure 5 contains the participant flow for the study. We began with 2,543 girls from the grade five enrollment lists. 2,459 were enrolled into the evaluation sample at the point of baseline and randomization; a girl is defined as enrolled into the sample if a parent/guardian provided consent for her inclusion in the evaluation. Of these girls, 2,399 were interviewed in the child survey at baseline.[8] Of the full sample of girls enrolled in the evaluation sample at baseline, 2,387 were interviewed in the child survey at endline, and 2,434 were observed in parent-reported data at endline. The implied attrition rate between baseline and endline is then 2.9 percent for the child-reported data, and 1 percent for parent-reported data; this is not a substantive factor in our analysis.

[8] There were 16 girls from one treatment school who were not surveyed at baseline; erroneous surveys were conducted with girls from another, adjacent school. This error was discovered following the first tracking survey. Data from the incorrect school was not included in analysis, as these girls were not part of the defined evaluation sample. The girls from the correct treatment school were then surveyed in the second tracking survey and the endline. We consider them as enrolled from the initiation of the evaluation, but they were not included in the baseline.

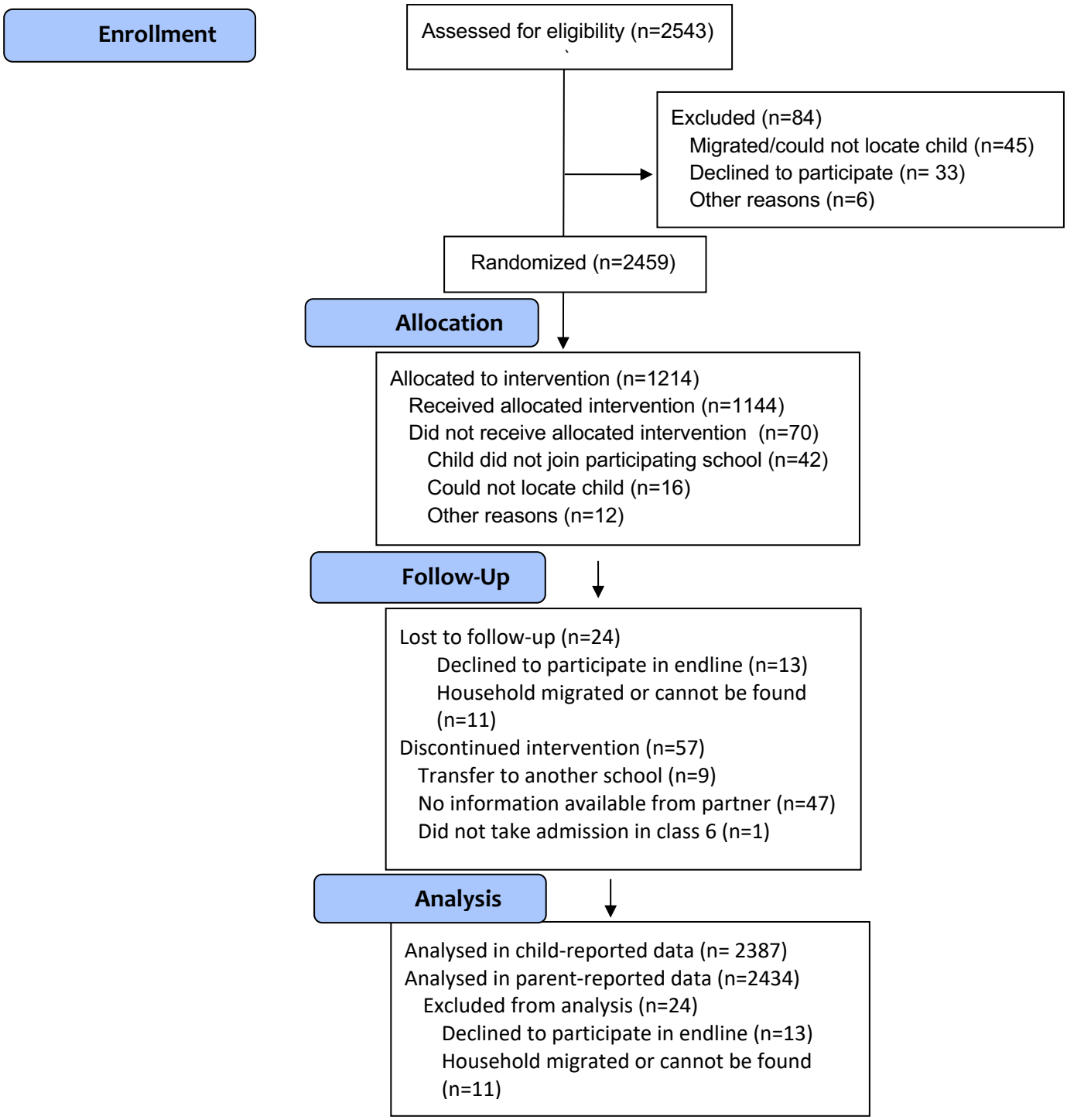

Figure 5: Flow Diagram - Child[9]

[9] Other Reasons for exclusion from the study include: illness or death of a child (2), parents declined to consent as child did not take admission in participating school (3), no one to provide consent (1). Other Reasons for not receiving allocated intervention include: child irregular or moved to different school (4), Child illness (2), Could not locate child (6).

3.2 Recruitment

As previously noted, households and girls were not recruited into this trial on an individual basis; all girls enrolled in schools that were identified as part of the evaluation sample were visited as part of the baseline survey and enrolled into the evaluation if and when they were surveyed. Recruitment was thus conducted during the baseline survey period, February 2, 2016 to July 16, 2016.

The first wave of follow-up was the first tracking survey, conducted between December 13, 2016 and January 25, 2017. The second tracking survey was conducted between December 12, 2017 and January 25, 2018. The endline survey was conducted between July 5, 2018 and January 2, 2019. The evaluation was designed to conclude following two full years of follow-up and concluded as projected.

3.3 Baseline data

The baseline data can be used to characterize sample demographics. 26% of the sampled households are members of a scheduled caste or scheduled tribe; 67% are members of a caste group denoted as OBC, or other backward caste, and the remainder are members of general caste households. 21% of households are Muslim, and 78% are Hindu. 22% of households reported primary dependence on self-employment in agriculture, and 8% on self-employment outside of agriculture. 53% reported dependence on wage employment, 1% on casual labor in agriculture, and 16% on casual labor outside of agriculture.[10]

Additional summary statistics at the household level are provided in Table 1. It should be noted that these summary statistics are parallel to those provided in the baseline report but have been updated to exclude a small number of girls (15) who were surveyed in error in a school that was erroneously identified as part of the sample. These girls were subsequently replaced in follow-up data collection with girls from the correct school.

The average household reported ownership of around six bighas of land, or approximately one hectare. 76% had an NREGA card, which reflects likely participation in the national guaranteed employment program. Households included around seven members and on average four children, of whom two were girls. Total consumption in the last month averaged around 26,000 rupees or $400.

Average summary statistics for the child outcomes mapped out in the evaluation design plan are provided in Table 2. Variable definitions are presented in the Evaluation design plan included as Appendix 7.2.

[10] The remainder of households report dependence on another economic activity not otherwise specified.

In sum, 3% of the sampled children had already dropped out of school at baseline. Attendance was relatively high. 89% attended school in the past week, and they attended 79% of the days the school was open. Our original lists of eligible students were based on grade 5 school rosters and 97 percent reported being in grade 5 at baseline. While the modal girl in our study is at an appropriate age for grade 5 (age 10-11), approximately a third of the sampled girls were older than that, indicating delays in their school progression. 17% of the sampled girls were married at baseline, though most married girls did not yet report cohabitating.

Reported child labor was prevalent among sample girls at baseline. 87% of respondents engaged in child labor during the twelve months before the baseline survey, and 64% of respondents were engaged in hazardous child labor. These child labor measures are based on the last year, but we felt that measures of hours worked could not reasonably be collected over such a long time horizon. We asked about hours worked through a time diary completed for a typical day during the previous work week. The average respondent reported one hour of economic activity and an additional 1.5 hours in the day in unpaid household services. At least in the week prior to the survey, on average children were not working full time.

We also report the summary statistics for a number of constructed indices of non-cognitive skills. These summary statistics are not directly interpretable but will be employed in the balance tests and to assess shifts in life skills over time.

Following the completion of the baseline, we also conducted balance tests to compare characteristics of the households in the treatment and control arms. The results are reported in Tables 3 and 4.

Table 3 reports the results of simple regressions in which the household characteristic of interest is regressed on a treatment indicator and strata dummies. We present the coefficient on the treatment indicator variable from each regression. We also present standard errors that are clustered at the school level, which is appropriate since this is the level of the randomization. Estimates indicate that households are more likely to be from other backward castes. We also estimate a seemingly unrelated regression specification that tests the joint null hypothesis that the treatment coefficient is equal to zero. Here, the p-value on this joint test of significance is 0.138, implying that we fail to reject the null hypothesis that the treatment coefficient is equal to zero at standard significance levels.

Table 4 reports similarly-constructed balance tests for child characteristics.[11] Here, the evidence suggests that girls in the treatment arm are characterized by somewhat more adverse circumstances at baseline as a number of our life skills indexes are negatively associated with treatment and differences, though imprecise, are large in magnitude. We also observe statistically significant higher levels of child work in the treatment group (more work, more

[11] This includes baseline engagement in child labor.

hours). The p-value on a joint test of significance of the treatment coefficient across all equations, constructed as described above, is 0.003, implying that we reject the hypothesis that child characteristics do not differ systematically based on treatment status. It will be important that these child characteristics are controlled for in our analysis as a robustness check.

3.4 Numbers analysed

The analysis sample includes all sampled girls who were represented in the endline survey. For the endline survey, training was conducted from June 18, 2018 to July 2, 2018, and field work was conducted from July 5, 2018 to January 2, 2019. The field team included 21 enumerators, three supervisors, two field monitors, two field managers, two back-checkers, and one project associate during the first month of data collection.

At endline, 2,387 child surveys were conducted; in addition, 2,358 household surveys were conducted. In the endline, there were 48 girls (in 47 households) for whom a household survey was conducted without a child survey. In 14 cases, consent was declined for the girl survey. 19 of these 48 girls had migrated away from their households permanently; two had migrated temporarily and had not returned by the point at which the survey concluded. Four child surveys were not completed due to the death of the child, and nine child surveys were not completed due to the child's mental disabilities.

There was also one girl surveyed whose household did not complete an endline survey; in this case, the head of household consented for the girl's participation but declined to complete the household survey.

24 girls in 23 different households attrited fully at endline with no data collection completed. In 10 cases, the household had migrated and could not be reached for follow-up. Consent was declined in 12 cases. In one case, a partial survey was completed but the household declined to continue due to limited time, and in one case, no information was available about the household's whereabouts.

If we examine patterns of attrition with respect to treatment arm, we observe the following. Among the 24 girls who fully attrited, 15 are from schools assigned to the control arm and 9 are from schools assigned to the treatment arm; the probability of full attrition is not significantly correlated with treatment, conditional on strata fixed effects (β=-.004, p=.362). Among the 72 girls who attrited from the girl survey, 45 are from schools assigned to the control arm, and 27 are from schools assigned to the treatment arm. The probability of attrition from the girl survey is lower in treatment schools, and this difference is significant at the 10 percent level (β=-.014, p=.087).

3.5 Outcomes and estimation

We discuss the estimated results relevant to the four key hypotheses below. For each family of outcomes, we present results corresponding to specifications (1) through (3) enumerated in section 2.8.1 above in Panels A, B and C, respectively, of Tables 5 through 8. In Panel A, we report the results of estimating specification (1), a simple specification that does not include baseline controls. In Panel B, we estimate specification (2), including controls for the baseline value of the outcomes of interest. In Panel C, we estimate specification (3) including controls for characteristics imbalanced at baseline. We note the number of missing observations only for those cases in which the number of observations is below the maximum number of child (2387) or household (2434) endline responses.

3.5.1 Hypothesis 1 Related Findings

To examine the impact of GEP on Hypothesis 1 outcomes, we have estimated the effect of treatment assignment on school dropout, grade progression, and two alternative measures of attendance. Coefficients are reported in Table 5. Results from Panel A of Table 5 are pictured in Figure 6 along with 95 percent confidence intervals.

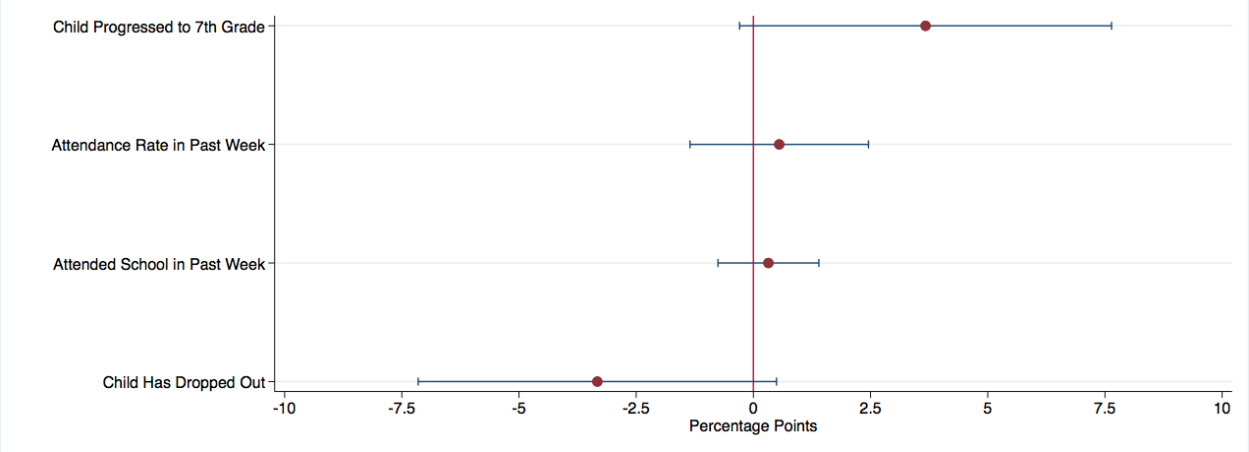

Figure 6: Treatment Effects Related to School Attendance

Our regression estimates in Panel A indicates that GEP reduced dropout by 3.3 percentage points, or 25% relative to a mean dropout rate of 13.2% in the control group. This estimate is statistically significant at the 10% confidence level, and the 95% confidence interval includes point estimates ranging from a 7.2 percentage point reduction in dropout to a 0.05 percentage point increase in dropout. The estimates in Panel B and C are of comparable magnitude, and the coefficient is significant at the 5 percent level in Panel C.

Table 5 also indicates that GEP correspondingly increased grade progression by 3.7 percentage points, or 4.3% relative to a mean grade progression rate of 86.5% in the control group. This estimate is similarly significant at the 10 percent level, and the 95% confidence interval includes

point estimates ranging from a 7.7 percentage point increase in grade progression to a 0.3 percentage point decrease in grade progression. The estimates in Panels B and C are similar in magnitude but slightly larger and are significant at the 5 percent level. Moreover, the fact that the estimated coefficients for dropout and grade progression are nearly identical suggests the effects of the program operate solely through reducing dropout and not via shifts in the probability of promotion to a new grade conditional on enrolling in school.

Turning to the two included attendance measures, 286 observations are missing for these endline attendance measures because of temporary school closures. The additional change in observations from the previous columns owes to dropouts. The regression coefficient based on the continuous attendance measure indicates that GEP increased attendance in the past week by less than a percentage point. Similarly, the regression coefficient based on the binary measure suggests an increase in attendance by 0.6 percentage points, and this estimate is similarly imprecise. Mean attendance in the control arm exceeds 90%. Overall, the main effect of the GEP on schooling seems to be on whether the girl stays and progresses through school with nothing measurable on the higher frequency attendance measures.

3.5.2 Hypothesis 2 Related Findings

To examine the impact of GEP on Hypothesis 2 outcomes, related to girls' life skills, we have estimated the effect of treatment assignment on four objective, task-based measures as well as 21 survey-based variables. For outcome variables that are dummy variables or measures of time and effort, we report the magnitudes in terms of percentage effects; this applies to variables capturing future discounting and the mirror drawing task, as well as the dummy variable for whether the girl is married. For outcome variables that are calculated as indices, however, we follow conventions in the literature to report the effect magnitudes in terms of standard deviations in the control arm; this applies to the scavenger hunt index as well as all of the non-cognitive indices constructed based on girl and parental reports.

Coefficients for the four objective measures are reported in Table 6-1; again, the results of estimating specifications (1) through (3) are reported in Panels A through C, respectively. With regards to the included objective measures, we see no evidence of statistically or economically significant treatment effects on the associated outcomes. For future discounting, 66.9% of control group girls prefer a smaller payment at present over a larger future payment, and treatment assignment does not appear to substantively alter this. Turning to the mirror drawing task, point estimates indicate only a marginal increase in effort exerted in response to treatment assignment. GEP assignment increased the number of attempted drawings by 0.06, or 1.7% relative to a mean of 3.3 drawings attempted in the control group. Similarly, the time spent on drawings increased only slightly for treated girls: GEP assignment increased the number of seconds spent on drawings by 2.2, or 1.8% relative to a mean of 119.5 seconds spent by girls in the control group. For the scavenger hunt completion index, we observe a 0.08 standard deviation decline in performance for girls assigned to treatment, but the 95% confidence interval

includes treatment effects ranging from a 0.2 standard deviation decline in performance to a 0.03 standard deviation improvement. In general, the magnitude of the estimated coefficients are consistent across panels.

In total, seven observations are missing from the analysis for both the future discounting and scavenger hunt measures, corresponding to the seven cases in which the respondent elected only to respond to the first section of the child survey. 70 observations are missing for time spent on mirror drawing measure, corresponding to the 70 respondents who did not attempt any mirror drawings.

The estimated results for survey-based measures of non-cognitive skills reported by the girl are reported in Table 6-2. Figure 7 is a visual representation of the results in Panel A of Table 6-2.

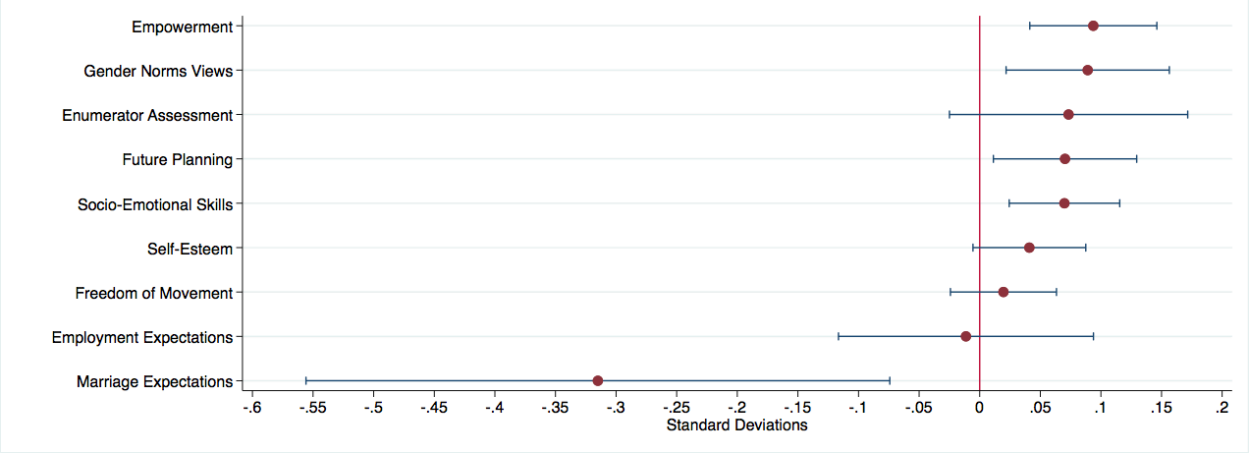

Figure 7: Treatment Effects Related to Child Expression of Life Skills

The results in Column (1) of Table 6-2 suggest a positive but imprecise effect of treatment on the likelihood that a girl is married, committed or engaged. The magnitude of this coefficient is large and merits discussion. GEP assignment is associated with a 4.2 percentage point increase in this outcome, or a 22.0% increase relative to a mean rate of marriage/engagement of 19.1% in the control group. This large difference is also imprecise. The 95% confidence interval includes values ranging from a 1.5 percentage point reduction in the marriage/engagement rates to a 9.9 percentage point increase. The coefficient's magnitude may be influenced by differences in the treatment and control group; when we control for differences in Panel C, the coefficient becomes negligible in magnitude (and still insignificant). For the single girl who was surveyed but her household was not surveyed, the child survey was used to estimate this endline measure.

We identify more sizable (and more precisely estimated) increases in response to treatment for a number of indices constructed based on girls' responses: GEP assignment increases the index of socio-emotional support by 0.07 standard deviations (significant at the 1% confidence level), increases the empowerment index by 0.09 standard deviations (significant at the 1% confidence

level), increases the index of future planning by .07 standard deviations (significant at the 5% confidence level), and increases the gender norms index by 0.09 standard deviations (significant at the 5% confidence level). These effects are consistent in magnitude and significance in Panels B and C as well. In addition, there is some evidence of a positive effect of the treatment on the enumerator assessment of the girl, an increase of .07-.10 standard deviations significant at the 10 percent level in specification (2) and 5 percent in (3) as reported in Panels B and C.

There are some less positive results in the table. GEP assignment reduces the average value of Cantril's ladder responses by 0.03, or 0.6% relative to a mean of 4.5 in the control group, indicative of weakly lower life satisfaction. The 95% confidence interval includes point estimates ranging from a 0.3 reduction to a 0.2 increase. Seven observations are missing for each of these endline indices, corresponding to the seven cases in which the respondent elected only to respond to the first section of the child survey (an eighth value is missing for the Cantril's ladder measure due to surveyor error).

The most concerning overall negative in the table comes from the marital expectations index. We measure a significant decline in response to treatment: GEP assignment reduces the marital expectations index by 0.3 standard deviations (significant at the 5% confidence level). 17 additional observations are missing for this index, corresponding to cases in which all component responses are missing because girls replied "Do Not Know" to associated questions. Part of the negativity of this index may come from the elevated marriage rate documented in column (1). That rate was spurious, but the negative effect on marriage expectations is consistent when estimated with controls in Panels B and C.

This negative change in marriage expectations may actually be a positive. The program discusses marriage and explicitly aims to help girls delay marriage and cohabitation. As such, all of the discussion around the problems of early marriage and the challenges girls face, discussion with the aim of helping prepare girls to manage those challenges, may induce girls to view that more negatively in the variables that feed into the index.

We do not see improvements in parental perceptions of life skills. The estimated results for indices constructed based on parental responses and new non-cognitive indices are reported in Table 6-3, Columns (1) through (10). For indices constructed based on parental responses in Columns (1)-(7), estimates are generally imprecise, heterogeneous in sign, and small in magnitude across alternative specifications. The one dimension along which we find significant effects is the index of parental perceptions of girl's strengths. This may also be a program positive as the program attempts to build the girl's understanding of herself and the constraints she is likely to face. A parent, interacting with a child as the issues come up, may easily assess the child's concern and attention as a negative.

Finally, the estimated coefficients for the new non-cognitive measures are reported in Table 6-3, Columns (8) through (10). Again, there is no evidence of any significant treatment effect on the

Rotter locus of control index, the perceived stress index, or the Rosenberg self-esteem index; the estimated coefficients are small in magnitude (less than 0.03 standard deviations) and statistically insignificant. They all point in a similar direction, as a negative locus of control implies more sense of agency, and similarly, a reduction in stress is welfare-enhancing. Seven observations are missing for each of these endline measures, corresponding to the seven cases in which the respondent elected only to respond to the first section of the child survey. Four observations are missing for parental perception of girl's self-efficacy, as the parent answered "Don't know" to all the relevant questions.

Figure 8 presents a visual summary of the findings from Table 6-3, reproducing the 10 confidence intervals and results for Panel A of the table.

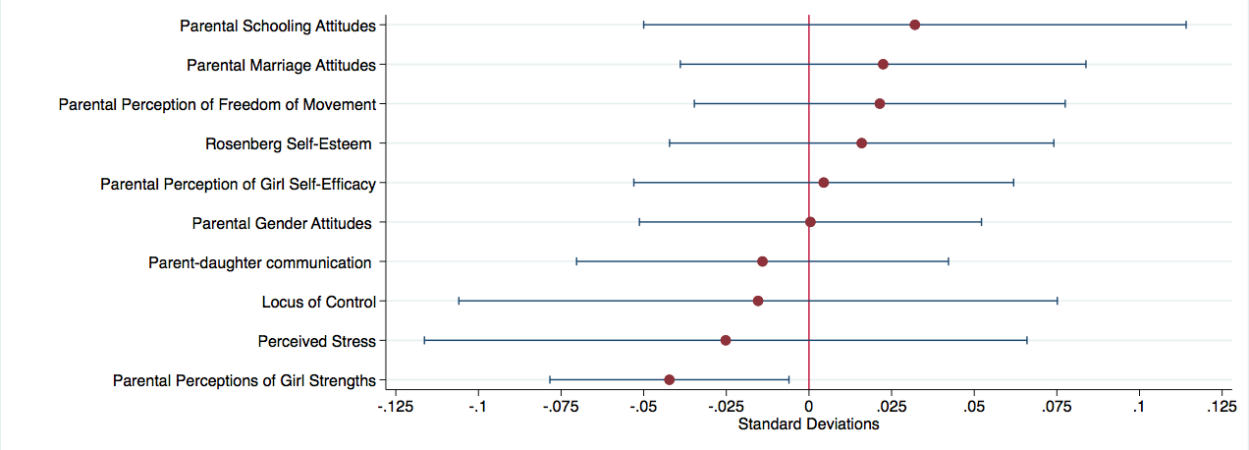

Figure 8: Treatment Effects Related to Parental Perceptions of Child Life Skills

3.5.3 Hypothesis 3 Related Findings

To examine the impact of GEP on Hypothesis 3 outcomes, related to child labor, we have estimated the effect of treatment assignment on ten survey-based measures that characterize extensive and intensive margin impacts on child labor. Coefficients are reported in Table 7; again, the results of estimating specifications (1) through (3) are reported in Panels A through C, respectively. We find impacts of treatment assignment that are not statistically significant at the 10% confidence level and are inconsistent in sign across specifications for these outcome measures. There does not appear to be any statistically significant or substantive change in child labor.

For outcome variables capturing the extensive margin – the probability of a child working for pay, working outside the home, engaging in child labor, engaging in hazardous child labor, or engaging in other worst forms of child labor – the estimated coefficients are uniformly small in magnitude (5 percentage points or smaller), variable in sign across specifications, and statistically insignificant.

Figure 9 presents Table 7 results for the extensive margin measures visually. It contains estimates of treatment effects and 95 percent confidence intervals for Panel A.

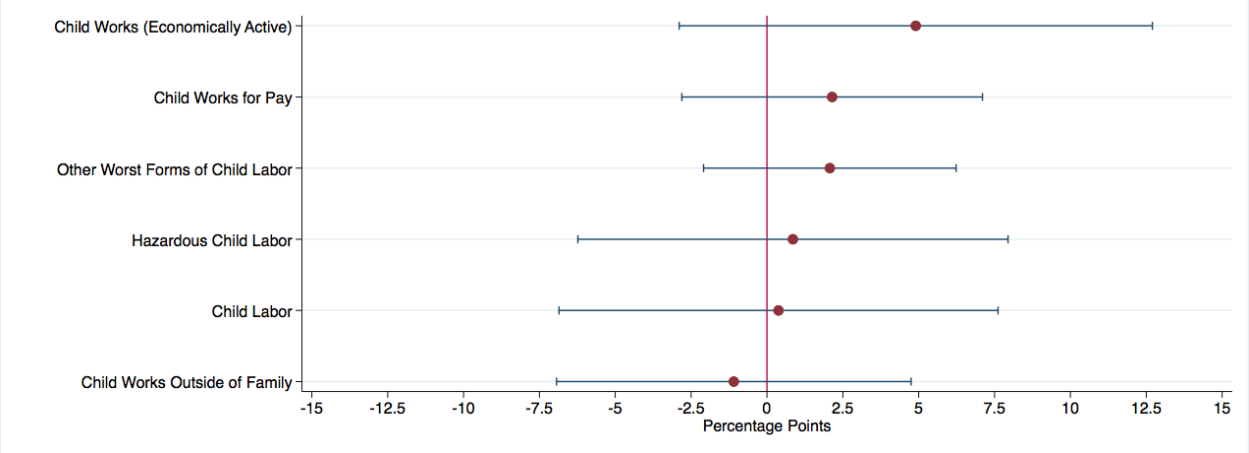

Figure 9: Treatment Effects Related to Child Labor Participation

Examining next the intensive margin measures of child labor, the estimates in Columns (7) through (10) of Panel A suggest small increases (less than 0.09 hours) in the number of hours worked in a day, the number of hours worked in unpaid work, the number of hours active, and the hours active outside the house; none of these estimates are statistically significant. These Panel A results for hours worked are presented visually in Figure 10 with 95 percent confidence intervals as well.

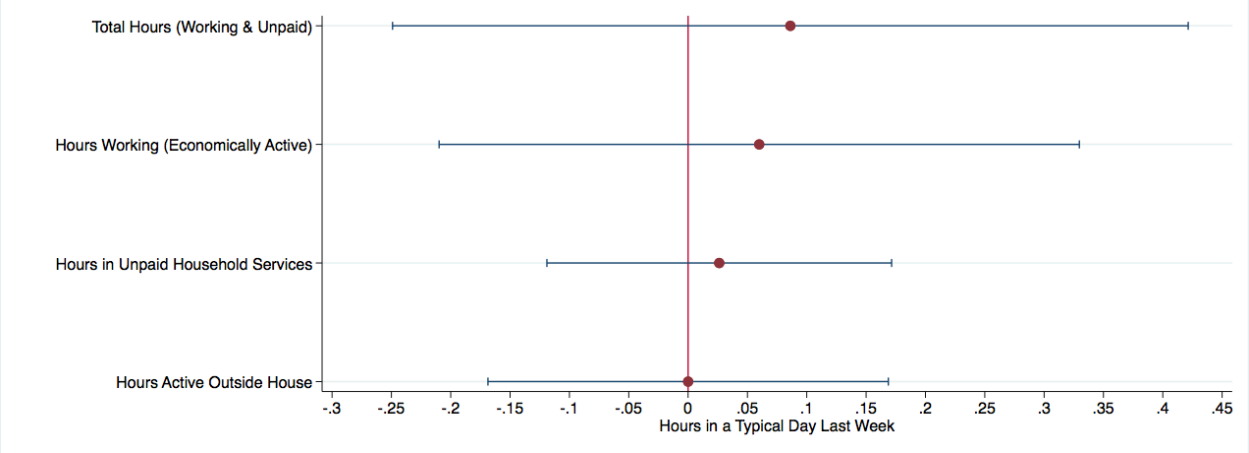

Figure 10: Treatment Effects Related to Child Labor Hours

In Panels B and C, the magnitude of these coefficients is reduced, and in some cases the sign is negative; again, none of the estimates are statistically significant. One observation is missing for each of these time allocation outcomes, corresponding to one respondent who did not answer question 311 in the child survey.

While child labor is somewhat imbalanced across the treatment and control groups at baseline, we reassuringly do not find that estimates change significantly with the inclusion of controls for lagged baseline measures. Given that a reduction in child labor represents a potential indirect consequence of GEP but not an explicit target, it is perhaps not surprising that impacts on this margin appear limited.

3.5.4 Hypothesis 4 Related Findings

To examine the impact of GEP on Hypothesis 4 outcomes, related to cognitive skills and academic achievement, we have estimated the effect of treatment assignment on five outcomes: time spent studying (outside of school), hours spent on school (in total), and three ASER cognitive test scores (for math, Hindi, and English). Coefficients are reported in Table 8; again, the results of estimating specifications (1) through (3) are reported in Panels A through C, respectively. Since cognitive tests were not conducted at baseline, for these outcome variables specification (2) includes controls for baseline school dropout status, attendance, grade progression, time spent studying, hours spent on school, and grades as reported in grade five.

With regards to time use, estimates are not statistically significant at conventional levels and are inconsistent in sign. Point estimates imply that GEP assignment decreases hours spent studying outside of school by 0.06 hours, or a decrease of 4.0% relative to the control group; however, the confidence interval includes changes ranging from a decrease of 0.2 hours to an increase of 0.1 hour. Conversely, treatment assignment increases hours devoted to school in total by 0.2 hours, or 2.5% relative to the control mean of 7.2 hours spent on school per day, but the 95% confidence interval ranges from a decrease of 0.2 hours to an increase of 0.6 hours. Neither of these estimates are statistically significant. One observation is missing for the time allocation variables, corresponding to one respondent who did not answer question 311 in the child survey.

Estimated test score effects are also imprecise, small in magnitude, and inconsistent in sign. Panel A results for ASER scores are presented visually in Figure 11 along with 95 percent confidence intervals.

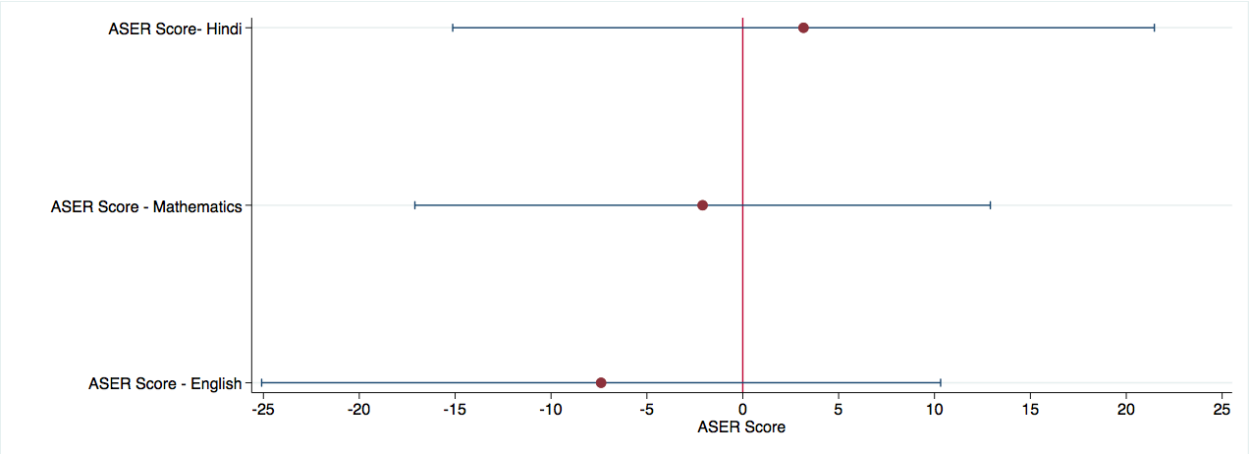

Figure 11: Treatment Effects Related to Test Scores

Estimates in Panel A suggest that GEP assignment decreases average math scores by 0.02 points, increases average Hindi scores by 0.03 points, and decreases average English scores by 0.07 compared to control scores of 2.35, 3.03, and 2.37 respectively. None of these estimates are statistically significant, and we also observe that the magnitudes are unstable in Panels B and C. 7 observations are missing for each of these endline test score measures, corresponding to the 7 cases in which the respondent elected only to respond to the first section of the child survey.

As for other outcomes, we only observe responses to GEP inputs over a two-year time horizon. Consequently, we cannot shed light on how these outcomes would be affected over the longer (seven-year) time horizon corresponding to the full-scale program.

3.6 Ancillary analyses

3.6.1 Pre-specified Heterogeneity Analysis

As described in Section 2.8.2, for all outcomes of interest, we pre-specified at baseline that we would conduct heterogeneity analysis based on school quality, child grade-for-age at baseline, and mother's education. School quality was assessed at randomization and is measured by an indicator that the child's school in grade 5 was above median. Because all targeted girls begin in the same grade, child grade for age is just age, and we examine heterogeneity by age. Mother's education is measured by an indicator for whether the mother has completed primary school. We also proposed to examine heterogeneous impacts based on whether the household had experienced the following types of household shocks: (1) economic shocks, (2) household illness or death, and (3) other shocks including crime and land/family disputes.

Our empirical approach interacts the source of heterogeneity with treatment. The interpretation of the coefficient on the treatment indicator is then the effect of treatment for the group where the heterogeneity variable is 0. The coefficient on the interaction gives the difference with

heterogeneity. Hence, the interpretation of the null hypothesis test on the interaction is testing the null that treatment effects do not vary with heterogeneity.

Across all six tables (Tables 9 – 12), there 308 statistical tests of no difference in treatment effects with heterogeneity, and only seven tests reject the null of no difference. We would expect 15 (false) rejections of the null even if all nulls are truly false. Hence, these seven significant results are consistent with all being false discoveries.

While they may be type 1 statistical errors, it is worth considering the seven significant results to examine if any patterns appear. They are:

- School attendance (conditional on not dropping out) is reduced by treatment when the household has experienced a crime shock (Table 9, column 3).
- School attendance (conditional on not dropping out) is increased by treatment when mothers have completed primary school (Table 9, column 3).
- Life satisfaction is lower with treatment in high quality schools than in low quality schools, although the combined effect is for no treatment effect on life satisfaction (Table 10-2, column 10)
- Parental perceptions of freedom of movement are lower with older children although the combined effect is for no treatment effect on freedom of movement (Table 10-3, column 3)
- Parent-daughter communication is lower in high quality schools than in low quality schools although the combined effect does not reject the null of no effect of treatment on parent-daughter communication (Table 10-3, column 4)
- The impact of treatment on work outside of the family is greater when households have been the victim or a death or illness (Table 11, column 3)
- The impact of treatment on the ASER English score is more negative when the household has experienced a crime shock (Table 12, column 5).

These findings do not appear to have any clear pattern. We conclude that there is little evidence to oppose the null that the impact of treatment does not vary with school quality, age, mother's education, or shocks.

3.6.2 Life skills

Room to Read's focus on life skills education comes from a detailed curriculum, and our endline survey contained 259 individual questions related to life skills measurement. While the analysis in Section 3.5.2 contains findings on the impact of treatment on pre-specified life skills aggregates, this section considers the impact of treatment on individual life skills questions and aggregates created based on Room to Read's current learning objectives.

Because we will be testing a large volume of hypotheses, it is important that our analysis be attentive to issues related to multiple hypothesis testing. P-values report the probability of seeing the observed deviation from zero under the null hypothesis that there is no effect of the treatment. P-values less than 0.05 are typically considered statistically significant. In our case, we will observe a p-value less than 0.05 at least once in every 20 regressions even if the null is true: i..e., even if there is no true treatment effect. To address this multiple hypothesis testing problem, we also report false discovery rate q-values. A q-value reports the fraction of null hypotheses with q-values at or below the q-value that are false rejections of the null hypothesis of no effect. Q-values depend on the number of null hypothesis tests, and in this case, we group all 259 individual questions plus nine indexes discussed below for a total of 268 regressions. Throughout, we present results that use the endline data only, conditioning on stratum fixed effects only and clustering standard errors by school. We define a treatment effect as statistically significant if it has a p-value less than 0.05 and a false discovery rate q-value of less than 0.25, or 25 percent.

We group our discussion below into two categories: individual questions and Room to Read targets. The individual questions section summarizes changes in life skills that we observe in individual questions. We relate these questions to the actual curriculum used in our life skills sessions as well as the summary results presented in Section 3.5.2. The Room to Read targets section groups individual questions, as appropriate, into the life skills targets designated explicitly by Room to Read. We also create an index for each of the life skills targets based on the associated individual questions and report the impact of treatment on each target.

3.6.2.1 Individual Questions

Table 13 contains our findings for the 259 individual questions asked regarding life skills in the endline survey. We have grouped them based on whether they are explicitly addressed in the grade six or grade seven curriculum or addressed only indirectly; we refer to these two question categories as “explicit” and “indirect”. It is important to bear in mind that the distinction between whether an item is explicitly addressed is based on whether we could identify an exact curricular match to the question; all of the life skills that we denote as not explicitly addressed may be indirectly addressed in the program. We have also grouped them by the magnitude of the effect of GEP on responses. A treatment effect of at least 15 percent is labeled “large”. This is arbitrary. Significant is defined as a p-value less than 0.05 and a q-statistics less than 0.25, and thus has a clear statistical interpretation.

Table 13 contains all the questions related to explicitly addressed subjects and any of the indirect questions that have effects that are observed to be significant and large. Indirect life skills related questions that are either not significant or not large are omitted from the table.

The general finding that girls seem to perceive improvements in social and emotional support and empowerment and that they answer with more positive gender norms and clear evidence of

increases in future planning is evident in the individual questions of Table 13. We see significant, large, taught effects on goal-setting and awareness of barriers girls face. Emotional regulation appears less successful.

For the indirectly-addressed skills, we observe large, significant effects on many questions related to the child expressing her own agency and influencing her own life. These are clear goals of the program even if the questions do not have an exact curricular match. We believe they can still be interpreted as an immediate consequence of the program rather than an ancillary outcome that follows from the program. The fact that the program enables girls to build life skills more broadly can be interpreted as a positive consequence of the intervention.

In addition, given that the evaluation detects effects on life skills that were not explicitly addressed in the curriculum, this minimizes the potential risk that the surveys are primarily detecting differential expression of life skills by girls who have been exposed to the programs (e.g., girls are informed that they should have higher aspirations, and thus they state that they have higher aspirations). Given that we see enhanced life skills across a range of dimensions that were not explicitly addressed, it seems that the observed effects do not simply reflect differential expression or reporting.

3.6.2.2 Room to Read Targets

The life-skills explicitly targeted by the GEP curriculum are *self-confidence, expressing & managing emotions, empathy, self-control, critical thinking, decision-making, perseverance, communication, creative problem solving*, and *relationship-building*. Our life skills-related questions are informative about nine of these life skills, with critical thinking left unevaluated.

Figure 12 pictures the impact of GEP on the nine target life skills. Point estimates and 95 percent confidence bounds are pictured for each. These life skills are standardized indexes (using the control mean and standard deviation) that aggregate the results of individual questions. Table 14 lists the questions that enter into each of the nine indexes and provides point estimates of the treatment effect on each individual question as well.

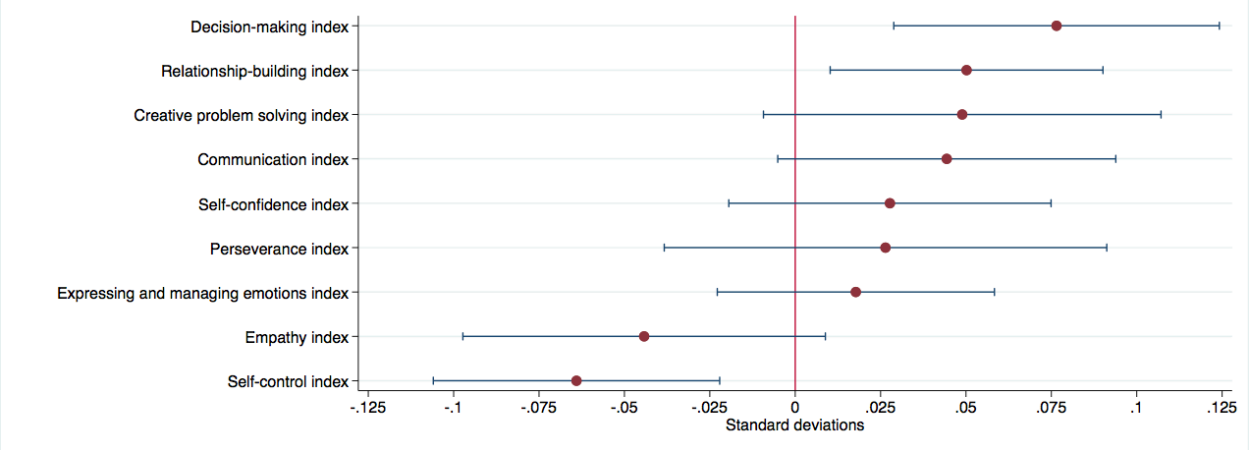

Figure 12: Impact of GEP on Life Skills Indexes

The largest effects of GEP are on the life skills related to decision-making. This large, positive effect seems to be driven by the child taking charge of her schooling and life-course decisions. We find a 17 percent increase in the child self-identifying as primarily in charge of her schooling decision. We observe greater than a 40 percent increase in the child deciding who and when she will marry (although we find no change in the probability the child is married by endline, approximately age 14). The largest change we observe is in the probability the child feels she will decide when she moves in with her in-laws. 3.5 percent of control girls feel they will have a say, while 6 percent of treatment girls feel they will have control over this decision. These magnitudes are small (girls in general express little agency over this choice), but the treatment's impact on their sense of agency is meaningful. Goal-setting is also impacted by the intervention, with a 20 percent increase in girls' ability to articulate a goal.

The next largest change is in relationship building. That impact seems driven by improvements in the child's ability to meet with friends outside of school. This mirrors some of the activities within the curriculum. It seems that classroom activities are generally translating into enhanced life skills as measured by these questions.

Creative-problem solving is another life skills dimension where GEP seems to have an effect. That effect is driven by the impact of the intervention on the girl's ability to articulate a strategy to communicate when someone does not understand her. This type of activity is practiced directly in the curriculum, and girls clearly feel more capable in that dimension.

This general pattern of curriculum translating into articulation of life skills also appears within communication. Girls clearly feel that they can communicate more effectively around marriage and future work decisions.

Interestingly, we observe declines in empathy and self-control associated with GEP. In both these cases, the negative impact of the intervention comes through parental assessments of children rather than from the child. They feel she is less likely to think before acting and has trouble forcing herself to pay attention in class (self-control). They also feel she is less helpful to others (empathy). These parental assessments may both reflect the child being more assertive about her own needs. As such, these decline in life skills with treatment may reflect some of the impact that the treatment is having on the girl's own agency over her life.

3.7 Harms

There is no evidence of any harm associated with program impacts related to school dropout, grade progression, or school attendance given that we find significant declines in dropout and increases in grade progression, along with positive but not statistically significant changes in attendance. Similarly, given estimated impacts on time in school, time studying, and cognitive test scores that are small in magnitude and inconsistent in sign, there is little evidence of any harm that would be induced by program impacts along these margins.

More generally, we do not identify statistically significant effects indicative of harms imposed on participants along most alternative margins. While Panel A point estimates imply that GEP assignment is associated with 4.2 percentage point increase in marriage/engagement rates, the associated change is only 1.1 percentage points when controls for imbalanced baseline measures are included in Panel C. In any case, neither estimated effect is statistically significant at conventional levels, and so we cannot reject that the net impact on marital status is zero. We interpret estimated changes along other margins, such as self-perceived stress and life satisfaction, similarly. We cannot reject that differences between those assigned to the treatment and control groups are zero.

One source of concern is that we do estimate a statistically significant decline in the marital expectations index, indicating that girls expect to marry earlier, move in earlier with their in-laws, and/or to have less say regarding future marriage decisions. This finding does indicate that program participation may induce harm along this margin to the extent that these changes in expectations cause stress or reflect increased future likelihood that girls will be forced to marry or to marry below the legal age of consent.

Finally, turning to measures of child work activity, we do not observe any statistically significant effects on any measures estimated. Thus, it appears that harm along these margins is unlikely.

4 Discussion

4.1 Limitations

This study evaluates a life skills education / mentoring program for adolescent girls in Ajmer, Rajasthan. The program is designed to begin with grade six and end with the completion of

secondary school (grade twelve). The evaluation takes place after the completion of grade seven, two years into the program.

While two years is an informative window, the program's ultimate objectives are to facilitate the completion of secondary school and to empower these girls, as they transition into adulthood, to take control of their lives. Hence, at two years, we are far removed evaluating the program's impact on its ultimate objectives.

How informative is a two-year window for the program's ultimate objectives? For schooling, completion of grades six and seven is a necessary condition for finishing secondary school. We observe that the intervention reduced dropout rates, relative to a rate of progression to seventh grade of 85.5% among control group girls. While we cannot say whether this translates into higher secondary school completion rates, staying in school is necessary for that to occur.

For life skills and cognitive performance, we are hesitant to extend beyond the period studied. Some characteristics take time to manifest. Hence, the lack of an impact of treatment on the task-based life skills measures may reflect that it takes more time than our window to observe the internalization of the survey-based non-cognitive effects we document. Likewise, it may take time for those survey-based measures to appear in cognitive performance. Alternatively, the survey-based measures may reflect what is taught in the classroom but not retained over a longer time horizon. Ultimately, our two-year time horizon is not informative about the long-term consequences of the program.

4.2 Generalizability

There are two important issues relevant to generalizability: geography and program.

On the geography side, Ajmer Rajasthan was not randomly chosen. It is a high growth area that is relatively accessible but lacking in similar programs. Rajasthan's Ministry of Education was an enthusiastic partner, and local schools were open and receptive to Room to Read's engagement.

Compared to elsewhere in Rajasthan, our population lives on smaller plots, has larger households, and contains more Muslim and Other Backward Caste households. Table 15 provides some simple summary statistics comparing variables measured in our baseline survey to similar variables measured in the 2007-08 District Level Health and Facility Survey. 68% of our sample is OBC, compared to 45% of Rajasthan or 33% of India. Similarly, 21% of our sample is Muslim, while only 10% of Rajasthan and 12% of India are Muslim. This relatively disadvantaged population seems to have higher school enrollments than Rajasthan in general, but that is partly an artifact of our survey following the comparison data by eight years in a setting with rising enrollments. Our qualitative research emphasized that this is an area where women

perceive few professional opportunities, although we do not know if that perception differs from elsewhere in Rajasthan.

One way to assess the generalizability of our findings to elsewhere is to consider the heterogeneity analysis in section 3.6. If a reader held a strong hypothesis that the impact of life skills education was invariant to background characteristics, there is nothing in our heterogeneity results in section 3.6 that would reject such a hypothesis. That said, we would hesitate to generalize these findings to other contexts.

Room to Read's life skills curriculum is thorough and intensive. Some efforts to teach life skills are more concentrated. For example, Ashraf et al (2018) focus on the impact of an after-school class that lasts two weeks and teaches one skill: negotiation. However, the general approach of Room to Read is widespread. The Life Skills Education program implemented by the Department of Public Instruction (DPI) Karnataka in the Bangalore area meets weekly for 12 to 20 sessions (Srikala and Kishore 2010). The Girls First program in Bihar worked with girls once a week for 23 weeks (Leventhal et al 2015). The Balika program in Bangladesh met weekly for 18 months (Amin et al 2016). The Compass program in Ethiopia met for 90 minutes weekly for 10 months (Stark et al 2018). Some are even more time intensive. The Ishraq program in rural Egypt met four times a week, three hours per meeting, for 30 months (Brady et al 2006). All of the above rely on educated young women to play a social mobilizer type role. One key difference in the Room to Read intervention is that the Room to Read program is designed to continue through secondary school while the other above programs have a more limited time horizon. Whether that influences our findings directly, two years into the program, is unclear.

4.3 Interpretation

Why did the life skills education combined with mentoring improve schooling progress and life skills?

The impact on education does not appear to come through attendance, as we see little impact of treatment on attendance rates. This is somewhat of a surprise given our qualitative research which highlighted many instances of girls being excited about school days when the life skills classes are held. One way to reconcile these narratives with our findings of no real impact on attendance is that the life skills days are only once every other week. Between their infrequency and the potential of other, random effects that might coincide with those days, girls might feel more enthused about attending schools on those days without any measurable effect in the data.

So why, then, do we observe improvements in schooling progression and declines in drop-out? In our qualitative work, interest in schooling emerged in part because girls began doing more of their schoolwork in groups. One student in Paner village remarked "Last year I got a C but this year I got a B." She explained that this happened "Because we fight less and sit and study together" as a result of the influence of the SM.

This idea of building a community of mutual support came up frequently in our qualitative work. One girl from Naka Madar, when asked what lessons she liked, remarked: “That we should help everybody. If someone is in trouble, we should ask them what is bothering them and help them. We should stay away from unknown people. If they touch you then you should go and tell somebody.”

If this mutual engagement and support is important in understanding our results for school progression, why don’t we see it in our test results? First, our test results are based on a test administered at home. Perhaps it maps poorly to school lessons. Second, there are many reasons to drop out beyond performance. It may be the case that some of the value of studying together comes from the *together* rather than what is being studied.

This idea of the intervention working through building a community is also supported by the child’s expression of life skills in reaction to survey questions. We observe improvements in the expression of socio-emotional characteristics and a sense of empowerment over their environment.

While we see little in the way of effects on parental perceptions, this theme of changing social interactions came out in the qualitative interviews with parents. One parent in Kayampura was very positive about the program. She remarked that: “(The SM) teaches them how to be obedient to parents, how to talk with them, not to fight at home, how to behave with friends and all. She teaches them manners on how to talk with other friends and outsiders.”

Self-regulation came up frequently in the qualitative work as a benefit of the program that helped girls build their community. One girl from Paner observed: “We learned how to live and work in communities, respect elders, to not be angry.” She learned a strategy to manage her anger: “Sometimes I get angry . . . didi told us to distract ourselves by reading a book or listening to music.”

A second main channel for improvements in schooling and expression of life skills may come from the program’s impact on goal setting. We find substantive effects of treatment on future planning, and we know from other studies such as Jensen (2010) that improvements in the child’s understanding of opportunities can increase schooling. Our qualitative findings emphasized improvements in goal setting. For example, one girl from ALS Nagar interviewed at baseline could only nod in the negative to a question about what she’d like to be when she grew up. By endline, she said that she wanted to be a police officer and could outline how she might reach that goal. Another girl from Paner was more explicit: “I didn’t know earlier what I wanted to do till I spoke with SM didi.” She wants to go to college in Roopangarh in order to become a teacher. Indeed, in baseline a frequent problem was that girls did not understand the meaning of “goal”. Having a goal and a plan to achieve it seems to be a clear outcome of the program.

One interesting note is that girls also appear to be more pessimistic about their marriage prospects as a result of the intervention. One danger in a life skills curriculum that draws attention to the challenges girls face is that girls would become discouraged. Despite rising pessimism in marriage prospects and no improvement in employment prospects, on net girls still seem to stay in school more and feel more empowered in their life. Hence, while this discouragement effect does not appear dominant in this intervention, it appears worth consideration in other life skills settings.

4.4 Policy Implications

In this evaluation, we utilize a randomized controlled trial to estimate the effects of the Girl's Education Program, an intervention built around a life skills curriculum implemented in school for girls in grades six and seven. Comparing girls in the treatment group to the control group, we find that the GEP reduced school dropout by 25 percent and improved grade progression by 4.3 percent.[12] These improvements in schooling come without evidence of meaningful improvements in school attendance or cognitive skills. These improvements in schooling are associated with improvements in the child's expression of life skills in survey questions. Girls perceive improvements in social and emotional support and empowerment. They also demonstrate more positive gender norms and clear evidence of increases in future planning. However, these improvements in life skills do not seem to be associated with substantive changes in child marriage or child labor.

Considerable emphasis has been placed recently on the economic and happiness values of soft skills (Deming 2017), and many researchers have emphasized the role of early life experiences in defining them (e.g. Heckman and Mosso 2014). Our findings highlight the malleability of the expression of life skills in adolescence.

Our setting is one where there are ample reasons to expect difficulties in impacting the expression of life skills. There are few professional opportunities for women, early marriage is common place, and the public visibility of women is low. Despite this setting, girls seem to absorb the lessons of the life skills curriculum and express them in response to questions. Hence, our findings present reasons to be optimistic about the possibility of teaching life skills in school to adolescent girls.

In our setting, girls seemed to enjoy the classes. They help build the sense of community and support around the girl and help her set goals, all of which may be helpful in keeping girls in school.

[12] As a comparison, this effect is similar in magnitude to the 32 percent increase in age-appropriate secondary school enrollment associated with a program that provided girls in Bihar who continued to secondary school with a bicycle (Muralidharan and Prakash 2013).

While our findings highlight that early life experiences are not entirely determinant in the expression of non-cognitive skills, it is important to bear in mind that there are many factors at play that influence a child's course in life. We found little in the way of effects on cognitive performance, and marriage and child labor are both positively correlated with treatment status. Indeed, in our qualitative work, interviewees frequently expressed that the life skill lessons and mentoring sessions had limited capacity to alter power relations within a family or change the influence of outside pressures. Hence, while the integration of life skills training into school curriculum seems promising for developing those skills, it is no panacea for the ills of poverty and underdevelopment.

Given that GEP continues throughout secondary school, future analyses of longer-term impacts on school completion, marriage rates and child labor will be helpful in complementing the conclusions we have reached based on this two-year time horizon.

5 Other information

5.1 Registration

This study was registered in the AEA RCT Registry, RCT ID AEARCTR-0001046, in 2016.

5.2 Protocol

The intervention was administered by Room to Read without interference from the research team, following Room to Read's protocol for the implementation of its Girls Education Program.

5.3 Role of the funder

Funding for the evaluation was provided by United States Department of Labor under Cooperative Agreement IL-26700-14-75-K-25 as well as a grant to Williams College from the Abdul Latif Jameel Poverty Action Lab. Room to Read funded the program through its usual budgetary process. This material does not necessarily reflect the views or policies of the United States Department of Labor, nor does the mention of trade names, commercial products, or organizations imply endorsement by the United States Government.

6 References

Amin, S., Ahmed, J., Saha, J., Hossain, I. and Haque, E., 2016. Delaying child marriage through community-based skills-development programs for girls. Results from a randomized controlled study in rural Bangladesh.

Ashraf, N., Bau, N., Low, C. and McGinn, K., 2018. Negotiating a better future: How interpersonal skills facilitate inter-generational investment.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), pp.289-300.

Brady, M., Assaad, R., Ibrahim, B.L., Salem, A., Salem, R. and Zibani, N., 2006. Providing new opportunities to adolescent girls in socially conservative settings: The Ishraq program in rural Upper Egypt.

Cantril, H. (1965). The pattern of human concerns. New Brunswick, NJ: Rutgers University Press.

Cohen, S., Kamarck, T., and Mermelstein, R. (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24, 386-396.

Coleman, M. and DeLeire, T., 2003. An economic model of locus of control and the human capital investment decision. The Journal of Human Resources, 38(3), pp. 701-721.

Deming, D.J., 2017. The growing importance of social skills in the labor market. *The Quarterly* Journal of Economics, *132*(4), pp.1593-1640.

Hamad, R., Fernald, L C., Karlan, D., and Zinman, J., 2008. Social and economic correlates of depressive symptoms and perceived stress in South African adults. Journal of Epidemiology and Community Health, 62(6), pp. 538-544.

Heckman, J.J. and Mosso, S., 2014. The economics of human development and social mobility. Annu. Rev. Econ., *6*(1), pp.689-733.

Heckman, J.J., Stixrud, J., and Urzua, S., 2006. The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. Journal of Labor Economics, *24*(3), pp.411-482.

Jensen, R., 2010. The (perceived) returns to education and the demand for schooling. *The* Quarterly Journal of Economics, *125*(2), pp.515-548.

Joshi, S. and Srivastava, R., 2009. Self-esteem and academic achievement of adolescents. Journal of the Indian Academy of Applied Psychology, 35, pp. 33-39.

Leventhal, K.S., Gillham, J., DeMaria, L., Andrew, G., Peabody, J. and Leventhal, S., 2015. Building psychosocial assets and wellbeing among adolescent girls: A randomized controlled trial. Journal of adolescence, 45, pp.284-295.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Rotter, Julian B (1966). "Generalized expectancies for internal versus external control of reinforcement". Psychological Monographs: General and Applied, 80: pp. 1–28.

Scales, P.C., Benson, P.L., Dershem, L., Fraher, K., Makonnen, R., Nazneen, S., Syvertsen, A.K. and Titus, S., 2013. Building developmental assets to empower adolescent girls in rural Bangladesh: evaluation of project Kishoree Kontha. Journal of Research on Adolescence, 23(1), pp.171-184.

Srikala, B. and Kishore, K.K., 2010. Empowering adolescents with life skills education in schools–School mental health program: Does it work?. Indian Journal of Psychiatry, *52*(4), p.344.

Stark, L., Seff, I., Assezenew, A., Eoomkham, J., Falb, K. and Ssewamala, F.M., 2018. Effects of a social empowerment intervention on economic vulnerability for adolescent refugee girls in Ethiopia. Journal of Adolescent Health, 62(1), pp.S15-S20.

7 Appendices

7.1 Overview of survey field practices

7.1.1 Baseline survey

The baseline survey was launched in February 2016 and began with a two-week training of enumerators, field supervisors, and back checkers. In total 26 staff were trained, of which 20 were recruited as enumerators, four as field supervisors, and two as back checkers. All the enumerators were female; the majority were recruited locally in Ajmer district following an advertisement and interview process, and were aged between approximately 25 and 40. The local enumerators were supplemented by three more experienced enumerators and one field supervisor who had been engaged in previous J-PAL projects in Bihar state and who were resident in Ajmer for the duration of the survey. The full team was further supplemented by two field monitors and managed by a field manager and our research assistant, Mohar Dey.

The training process focused on developing enumerator skills. Key points included strategies to locate respondents within the community; the importance of informed consent and how to correctly structure the consent process; establishing a rapport with respondents as well as with other stakeholders in the community; maintaining fidelity to the questionnaire; full comprehension of the questionnaires themselves; and correct use of the tablets. (All data collection was implemented using ODK software on handheld tablets.)

The survey teams then deployed to the field using household rosters that were constructed based on the lists of enrolled girls obtained from sampled schools. The information provided by the schools typically included the name of the head of household and the child herself, as well as some identifying information about the location of the household. In general, however, it was

also necessary for enumerators and field supervisors to work with community members to locate each household. Field supervisors and field managers would also make courtesy visits to community stakeholders (including the sarpanch or village leader, school headmaster, and teachers) when they first arrived in the community in order to introduce the team and outline the survey's objectives.

Each survey included a minimum of two visits to the household, as the survey administered to the girl herself was divided into two parts. This choice was made in order to maximize attention and avoid fatigue; in addition, the first visit was used to introduce a scavenger hunt task to the girl, so that she could engage in the scavenger hunt prior to the second visit. However, many households required more than two visits total to complete the data collection process, particularly as the household survey included multiple modules to be answered by different individuals. (For example, introductory modules including household rosters were administered to the head of household or the individual most knowledgeable about the household. Modules collecting information about perception of the child's life skills were administered to the individual primarily responsible for the child's care.)

The survey encountered two primary challenges. The first was retaining enumerators. Given that the survey team encompassed primarily younger women, there was a high degree of turnover as enumerators pursued other educational or professional opportunities. Retention challenges also increased in the latter part of the survey period, due to the intensifying heat. Of the 20 enumerators and four field supervisors recruited at the start of baseline, only eight enumerators and two supervisors continued until the completion of baseline surveying. Despite this challenge, however, we were able to maintain a survey team at the target size (around 20 enumerators and four supervisors) by recruiting and training new enumerators in waves.[13]

The second challenge encountered was community perceptions of the survey, particularly the surveys administered to girls and the use of the scavenger hunt. While household surveys are not unusual in this area, many households expressed surprise or discomfort when enumerators sought to survey their female children. In some cases, even administering surveys to adult females in the household seemed controversial. At times, other households in the community also reacted negatively to the presence of the survey team, and were suspicious of girls engaging in the scavenger hunt; in one community, there were rumors that the scavenger hunt was linked to black magic. Our field team also observed that these suspicions were sometimes heightened by differences in caste and socioeconomic background between the enumerators and the households they were visiting.

[13] In the first wave of supplementary hiring, six enumerators were hired, and four continued until the end of the survey. In the second wave of supplementary hiring, four additional enumerators were hired, all of whom continued to work until the conclusion of the survey. In the third wave of supplementary hiring, again four enumerators were hired and all continued to work until the conclusion of the survey.

We addressed these challenges in a number of ways. First and most importantly, we benefited from an excellent team of field supervisors and managers, the majority of which had experience conducting surveys in this area. They were able to train the enumerators in liaising effectively with households, reiterating the objectives of the survey and addressing any doubts or suspicions. Second, we maintained active relationships with community stakeholders. We regularly visited village leaders, school principals, and head teachers, and would also engage them in jointly speaking to households in which suspicions were expressed. Our objective always was to clarify that while households or children were free to decline to participate, the goal of the survey was only to collect information that would be kept confidential and that the risk for respondents was minimal.

For 44 girls included in the sample lists provided by the schools, we were ultimately able to conduct a household survey (including a survey of the child's parent or primary guardian) but did not collect a child survey. These 44 girls live in 43 separate households. Household surveys were collected without corresponding child surveys when the parent or child declined consent for the child's participation, particularly due to suspicions about the scavenger hunt (21 girls); and/or the child was residing in another community (two girls) or away from home for a long period of time during the summer vacation (21 girls).

For 16 girls included in the sample lists, living in 15 separate households, we were ultimately able only to conduct a survey of the child and did not complete a household survey. In five of these cases, the household head and other responsible adults were not available (i.e., working at another location, or working extremely long hours). In 10 of these cases, consent for the child survey was provided by an individual other than the household head; this was the primary caretaker of the child, either the mother or the grandmother. Subsequently, when the household head learned about the scavenger hunt, he/she declined to participate in the survey. In the remainder, the parent declined to consent to be surveyed.

In addition, 16 girls living in 14 separate households from one primary school (Arain) were omitted from the baseline in error. A different set of students enrolled in a different, adjacent primary school that is outside the evaluation sample were surveyed in their place. Given that these girls were not intended for inclusion in the sample, their data was subsequently dropped, and the correct set of girls were surveyed from the first tracking survey forward.

7.1.2 First tracking survey

The tracking survey was designed as a short survey, administered only to the sampled girls in the community who were interviewed at baseline. The objective was to collect information on girls' enrollment and attendance in school as well as basic information on girls' time allocation. If a girl in the sample was not found at her home address, detailed information was collected about her current location. If the family was not resident in the village, the field team endeavored to collect as much information as possible about their current residence and migration history, but

these short surveys were not meant to follow children to their new location in the event of their move.

Training for the first tracking survey was conducted from December 5, 2016 to December 9, 2016, and field work was conducted from December 13, 2016 to January 25, 2017. The field team included 19 enumerators, four supervisors, two field monitors and one field manager.

For the first tracking survey, the effective tracking sample excluded the girls from Arain, who were surveyed in error during the baseline. Thus the effective tracking sample was 2443 girls. 98% of the original sample of girls were interviewed in the first tracking survey. The reasons for attrition are summarized in the table below.

Appendix Table 1: Attrition in the First Tracking Survey

	Number of cases
Complete	2394
Household not found	4
Child deceased	2
Child not competent	2
No consent	10
Household migrated	14
Household migrated temporarily	2
Child migrated	15

In four cases, we could not locate the household and no information was provided by neighbors or other local informants. In those four cases, we also attempted to track the child down through the school she was originally attending, but no information was available

Four other subjects were also lost from the original sample, two because of death and two because the child was not competent to be interviewed. These two children deemed incompetent to be interviewed had likewise not been surveyed at baseline, though the parent was surveyed.

In ten cases the parent declined to provide consent for the child to be surveyed. (This included eight cases where the child had been successfully surveyed at baseline, and two cases in which even at baseline consent was declined for the child, but the parent did participate.)

Permanent or temporary migration prevented us from interviewing 31 children. We conducted re-visits in the case of temporary migration to attempt to find the child at home; the two cases of temporary migration that remain were cases in which the child had not yet returned when the survey concluded.

7.1.3 Second tracking survey

Training was conducted from December 4, 2017 to December 9, 2017, and field work was conducted from December 12, 2017 to January 25, 2018. The field team included 21 enumerators, four supervisors, one field monitors and two field managers.

94% of the original sample of girls were interviewed in the tracking survey. The reasons for attrition are summarized in the table below.

Appendix Table 2: Attrition in Second Tracking Survey

	Number of cases
Complete	2318
Household not found	9
No one was at home	5
Child not competent	6
Child deceased	3
No consent	16
Household migrated	26
Child migrated	54
Child migrated temporarily	22

In nine cases (compared to four in the first tracking survey), the household could not be located and no information could be provided by neighbors or other local informants. In four cases, no one was located at home during the duration of the survey. In 16 cases (compared to 10 in the first tracking survey), the parent declined to provide consent for the child to be surveyed.

Finally, the number of household migrations and child migrations (both permanent and temporary) increased since the first round of the tracking survey; we hypothesize that the increased number of child migrations may reflect partially the older age of the girls, and thus an increased probability that they are allowed to reside or travel away from home for longer periods. The temporary migrants who were not surveyed had not yet returned.

Table 1: Summary statistics for sampled households

	Mean	Standard deviation	Observations
Number of Study Subjects in Household	1.062	0.246	2427
Number of household members	6.838	2.811	2427
Number of boys in household (under 18)	1.379	1.026	2427
Number of girls in household (under 18)	2.438	1.359	2427
Other backward castes household	0.674	0.469	2427
Primary household source of employment = wage / salary earning	0.532	0.499	2427
Primary household source of employment = Self Employment Agirculture	0.215	0.411	2427
Primary household source of employment = Self Employment Nonagriculture	0.080	0.272	2427
Primary household source of employment = Casual Labor in Agirculture	0.013	0.114	2427
Primary household source of employment = Casual Labor in Non-Agirculture	0.157	0.363	2427
Non-food expenditures in Rupees (last 30 days)	9907	40338	2427
Food expenditures in Rupees (last 30 days)	16029	204568	2427
Durables expenditures in Rupees (last year)	123377	960822	2427
Land owned (bighas) [1]	6.283	15.959	1930
Land cultivated (bighas) [2]	2.301	12.525	1633
Household holds NREGA card	0.756	0.430	2427
Economic Shock [3]	0.606	0.489	2427
Crime Schock [4]	0.132	0.338	2427
Death / illness shock [5]	0.406	0.491	2427

[1] 7% of households, or 182 households, report that they own no land individually but access collectively owned land. 315 households, or 13%, cannot estimate the amount of land owned.

[2] 8% of households, or 206 households, do not report land cultivated because it is cultivated collectively, and in this case an additional 588 households (or 24%) cannot estimate the amount of land cultivated.

[3] Loss of employment or lowered income of any household member or bankruptcy of family business in last 12 months

[4] Criminal act including robbery, assault, or physically aggression; land dispute; or family dispute in last 12 months

[5] Death, serious illness, or accident of a household member in last 12 months

Note: Households with multiple study subjects occur as multiple observations. 16 study subjects completed a baseline child survey but no baseline household survey and thus are not represented in these baseline summary statistics.

Table 2: Summary statistics for sampled children

	Mean	Standard deviation	Observations
Stratification (Baseline school characteristics):			
Below median school quality	0.509	0.500	2407
Above median school quality	0.491	0.500	2407
Subject characteristics:			
Age	10.989	1.425	2419
Maternal Education (1=completed primary or above) [1]	0.172	0.377	2426
Girl's Marital Status (1=Married)	0.167	0.373	2421
Child has dropped out of school	0.025	0.156	2440
Child is in grade 6	0.975	0.156	2399
Any Attendance in last week (conditional on not dropping out) [2]	0.889	0.314	2026
Attendance Rate in last week (conditional on attendance) [3]	0.788	0.337	2026
Delay discounting	0.178	0.383	2399
Completed Mirror Drawings	2.396	1.248	2415
Mirror Drawings (seconds)	68.407	70.266	2399
Scavenger Hunt Index	-0.023	0.969	2398
Socio-emotional Index	0.017	0.464	2399
Freedom of Movement Index	-0.001	0.602	2399
Empowerment Index	-0.004	0.416	2399
Self-Esteem Index	0.021	0.481	2399
Future Planning Index	0.051	0.601	2399
Marital Expectations Index	-0.496	1.435	2399
Employment Expectations Index	-0.017	0.794	2399
Gender Norms Index	-0.003	0.509	2399
Cantrill ladder	7.955	2.418	2399
Enumerator Assessment Index	-0.027	0.891	2399
Parental Perception of Girl's Strengths	0.004	0.365	2425
Parental Perception of Girl's Self-Efficacy	0.024	0.623	2425
Parental Perception Freedom of Movement	-0.021	0.590	2425
Parent-Daughter Communication	0.002	0.422	2443
Parental Gender Attitudes	0.001	0.432	2425
Parental Schooling Attitudes	0.007	0.695	2427
Parental Marriage Attitudes	-0.005	0.516	2425
Child Works	0.914	0.280	2398
Child Works for Pay	0.844	0.363	2398
Child Works outside of Family Activity	0.697	0.460	2399
Child Labor	0.874	0.332	2398
Hazardous Child Labor	0.642	0.479	2397
Other Worst Forms of Child Labor	0.225	0.418	2399
Hours Economically Active in a day	1.052	1.691	2397
Hours in Unpaid Household Services in a day	1.447	1.448	2397
Total Hours Active	2.499	2.308	2397
Hours active outside house	0.825	1.454	2397
Hours studying at home	0.704	0.955	2397
Total hours spent on school	6.105	2.823	2397

[1] From endline survey: missing if child is not present in endline survey

[2] Missing if child has dropped out or her school was not open in past week

[3] Missing if child has dropped out, her school was not open in past week, or she did not attend school in past week

Table 3: Balance tests for household variables

	Control		Treatment		Difference	
	Mean	Standard dev.	Mean	Standard dev.	Coef.	(se)
Number of Study Subjects in Household	1.057	0.241	1.067	0.251	0.011	(0.015)
Number of household members	6.893	2.763	6.781	2.860	-0.111	(0.142)
Number of boys in household (under 18)	1.358	1.047	1.402	1.003	0.044	(0.051)
Number of girls in household (under 18)	2.456	1.340	2.419	1.380	-0.037	(0.063)
Other backward castes household	0.631	0.483	0.720	0.449	0.089**	(0.038)
Primary household source of employment = wage / salary earning	0.536	0.499	0.527	0.499	-0.009	(0.031)
Primary household source of employment = Self Employment Agirculture	0.210	0.407	0.220	0.415	0.010	(0.032)
Primary household source of employment = Self Employment Nonagriculture	0.072	0.258	0.089	0.285	0.017	(0.015)
Primary household source of employment = Casual Labor in Agirculture	0.015	0.120	0.012	0.108	-0.003	(0.005)
Primary household source of employment = Casual Labor in Non-Agirculture	0.162	0.369	0.151	0.358	-0.012	(0.019)
Non-food expenditures in Rupees (last 30 days)	10342	53692	9454	17907	-888	(1,700)
Food expenditures in Rupees (last 30 days)	21695	286167	10130	10954	-11565	(7,938)
Durables expenditures in Rupees (last year)	107787	549816	139609	1252968	31823	(40,099)
Land owned (bighas)	5.653	11.828	6.901	19.153	1.249	(1.108)
Land cultivated (bighas)	2.069	8.285	2.540	15.738	0.472	(0.763)
Household holds NREGA card	0.712	0.453	0.802	0.399	0.090	(0.063)
Economic Shock [1]	0.593	0.491	0.620	0.486	0.027	(0.024)
Crime Schock [2]	0.126	0.332	0.138	0.345	0.012	(0.017)
Death / illness shock [3]	0.396	0.489	0.417	0.493	0.021	(0.021)

The columns under the header "Difference" report the result of the regression of the row variable on an indicator for treatment and stratification fixed effects. The column labeled "coef" reports that coefficient on the treatment indicator which is the same as the difference between the treatment and control means. The column "(se)" contains the standard error on that coefficient, corrected for clustering on school. One household did not complete a roster and thus is not represented in these baseline summary statistics. * p<0.10, ** p<0.05, *** p<0.01

[1] Loss of employment or lowered income of any household member or bankruptcy of family business in last 12 months

[2] Criminal act including robbery, assault, or physically aggression; land dispute; or family dispute in last 12 months

[3] Death, serious illness, or accident of a household member in last 12 months

Table 4: Balance tests for child variables

	Control		Treatment		Difference	
	Mean	Standard dev.	Mean	Standard dev.	Coef.	(se)
Stratification (Baseline school characteristics):						
Below median school quality	0.515	0.500	0.502	0.500	-0.013	(0.096)
Above median school quality	0.485	0.500	0.498	0.500	0.013	(0.096)
Subject characteristics:						
Age	10.960	1.411	11.019	1.440	0.058	(0.082)
Maternal Education (1 = has completed primary school or higher) [1]	0.187	0.390	0.156	0.363	-0.031	(0.024)
Girl's Marital Status (1=Married)	0.141	0.348	0.194	0.396	0.053*	(0.028)
Child has dropped out of school	0.024	0.154	0.026	0.159	0.002	(0.010)
Child is in grade 6	0.975	0.155	0.974	0.158	-0.001	(0.010)
Any Attendance in last week (conditional on not dropping out) [2]	0.870	0.336	0.908	0.290	0.037	(0.025)
Attendance Rate in last week (conditional on attendance) [3]	0.768	0.353	0.808	0.319	0.040	(0.028)
Delay discounting	0.171	0.376	0.186	0.390	0.016	(0.027)
Completed Mirror Drawings	2.489	1.228	2.331	1.242	-0.158	(0.118)
Mirror Drawings (seconds)	69.452	65.875	67.318	74.578	-2.133	(6.092)
Scavenger Hunt Index	0.000	0.965	-0.048	0.973	-0.048	(0.077)
Socio-emotional Index	0.000	0.480	0.035	0.447	0.035	(0.030)
Freedom of Movement Index	0.000	0.577	-0.002	0.628	-0.002	(0.046)
Empowerment Index	0.000	0.406	-0.008	0.426	-0.008	(0.029)
Self-Esteem Index	0.000	0.497	0.044	0.462	0.044	(0.028)
Future Planning Index	0.020	0.610	0.084	0.589	0.064*	(0.034)
Marital Expectations Index	-0.401	1.348	-0.595	1.514	-0.194*	(0.105)
Employment Expectations Index	-0.002	0.788	-0.033	0.801	-0.030	(0.053)
Gender Norms Index	0.000	0.498	-0.005	0.521	-0.005	(0.034)
Cantrill ladder	8.029	2.395	7.877	2.440	-0.152	(0.152)
Enumerator Assessment Index	0.000	0.859	-0.056	0.922	-0.056	(0.052)
Parental Perception of Girl's Strengths	0.000	0.374	0.007	0.357	0.007	(0.033)

Parental Perception of Girl's Self-Efficacy	0.000	0.611	0.048	0.636	0.047	(0.038)
Parental Perception Freedom of Movement	0.000	0.532	-0.043	0.643	-0.043	(0.035)
Parent-Daughter Communication	0.001	0.415	0.002	0.429	0.001	(0.029)
Parental Gender Attitudes	0.000	0.424	0.003	0.439	0.003	(0.025)
Parental Schooling Attitudes	0.003	0.682	0.012	0.709	0.010	(0.051)
Parental Marriage Attitudes	-0.005	0.503	-0.005	0.530	-0.000	(0.033)
Child Works	0.884	0.320	0.945	0.227	0.061***	(0.021)
Child Works for Pay	0.829	0.376	0.859	0.349	0.029	(0.026)
Child Works outside of Family Activity	0.674	0.469	0.721	0.449	0.047	(0.035)
Child Labor	0.855	0.352	0.893	0.310	0.037	(0.024)
Hazardous Child Labor	0.620	0.486	0.665	0.472	0.045	(0.037)
Other Worst Forms of Child Labor	0.219	0.414	0.231	0.422	0.013	(0.027)
Hours Economically Active in a day	0.945	1.636	1.164	1.741	0.219*	(0.128)
Hours in Unpaid Household Services in a day	1.415	1.454	1.480	1.441	0.065	(0.070)
Total Hours Active	2.360	2.243	2.644	2.367	0.284*	(0.158)
Hours active outside house	0.719	1.387	0.935	1.513	0.216**	(0.099)
Hours studying at home	0.713	0.966	0.694	0.944	-0.018	(0.063)
Total hours spent on school	6.014	2.845	6.199	2.799	0.185	(0.274)

The columns under the header "Difference" report the result of the regression of the row variable on an indicator for treatment and stratification fixed effects. The column labeled "coef" reports that coefficient on the treatment indicator which is the same as the difference between the treatment and control means. The column "(se)" contains the standard error on that coefficient, corrected for clustering on school. One household did not complete a roster and thus is not represented in these baseline summary statistics. * $p<0.10$, ** $p<0.05$, *** $p<0.01$

[1] From endline survey: missing if child is not present in endline survey

[2] Missing if child has dropped out or her school was not open in past week

[3] Missing if child has dropped out, her school was not open in past week, or she did not attend school in past week

Table 5: School progression and completion

	(1) Whether child has dropped out	(2) Whether child progressed to 7th grade	(3) Attendance rate	(4) Attendance - dummy
Panel A: No controls				
Treatment	-0.033*	0.037*	0.006	0.003
	(0.020)	(0.020)	(0.010)	(0.005)
Observations	2433	2387	2089	2089
Adjusted R-squared	0.002	0.003	0.001	0.000
Panel B: Controls for baseline values				
Treatment	-0.035*	0.038**	0.004	0.003
	(0.018)	(0.018)	(0.009)	(0.006)
Observations	2433	2387	2089	2089
Adjusted R-squared	0.121	0.120	0.012	0.004
Panel C: Additional controls				
Treatment	-0.041**	0.042**	0.002	0.003
	(0.018)	(0.019)	(0.009)	(0.006)
Observations	2433	2387	2089	2089
Adjusted R-squared	0.142	0.134	0.015	0.006
Mean control group	0.132	0.865	0.918	0.982

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables. Column 1 uses child and household survey data. Columns 2-4 use child survey only. Columns 3 - 4 are conditional on school being open and child not having dropped out of school.

* p<0.10, **p<0.5, *** p<0.01

Table 6-1: Non-cognitive skills - Objective measures

	(1) Delay discounting	(2) Completed Mirror Drawings	(3) Mirror Drawings (seconds)	(4) Scavenger Hunt Index
Panel A: No controls				
Treatment	-0.000	0.056	2.172	-0.079
	(0.032)	(0.085)	(4.472)	(0.057)
Observations	2380	2387	2317	2380
Adjusted R-squared	0.004	0.002	-0.000	0.002
Q-statistic	0.991	0.827	0.827	0.428
Panel B: Controls for baseline values				
Treatment	-0.003	0.070	2.610	-0.072
	(0.032)	(0.085)	(4.535)	(0.055)
Observations	2380	2387	2317	2380
Adjusted R-squared	0.007	0.010	0.004	0.051
Q-statistic	0.997	0.790	0.833	0.532
Panel C: Additional controls				
Treatment	0.003	0.072	2.720	-0.065
	(0.032)	(0.082)	(4.559)	(0.055)
Observations	2380	2387	2317	2380
Adjusted R-squared	0.010	0.015	0.010	0.053
Q-statistic	0.915	0.737	0.833	0.617
Mean control group	0.331	3.269	119.5	0.000

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables. Q-statistics are False Discovery Rate corrected q-values based on Benjamini and Hochberg (1995). These are computed pooling all Table 6s results within a panel.

* p<0.10, **p<0.5, *** p<0.01

Table 6-2: Non-cognitive skills - Survey measures

	(1) Married	(2) Socio-emotional Index	(3) Freedom of Movement Index	(4) Empowerment Index	(5) Self-Esteem Index	(6) Future Planning Index	(7) Marital Expectations Index	(8) Employment Expectations Index	(9) Gender Norms Index	(10) Cantril ladder	(11) Enumerator Assessment Index
Panel A: No controls											
Treatment	0.042	0.070***	0.020	0.094***	0.041*	0.070**	-0.315**	-0.011	0.089**	-0.026	0.073
	(0.029)	(0.023)	(0.022)	(0.027)	(0.024)	(0.030)	(0.123)	(0.054)	(0.034)	(0.133)	(0.050)
Observations	2435	2380	2380	2380	2380	2380	2380	2380	2380	2380	2380
Adjusted R-squared	0.004	0.006	-0.000	0.011	0.001	0.003	0.009	0.002	0.008	-0.001	0.001
Q-statistic	0.415	0.041	0.827	0.017	0.311	0.100	0.073	0.956	0.073	0.956	0.415
Panel B: Baseline controls											
Treatment	0.011	0.063***	0.022	0.097***	0.034	0.065**	-0.198**	-0.000	0.088***	0.001	0.092*
	(0.018)	(0.023)	(0.023)	(0.027)	(0.023)	(0.031)	(0.081)	(0.039)	(0.034)	(0.131)	(0.047)
Observations	2435	2380	2380	2380	2380	2380	2380	2380	2380	2380	2380
Adjusted R-squared	0.326	0.023	0.006	0.018	0.044	0.020	0.298	0.190	0.020	0.007	0.041
Q-statistic	0.810	0.080	0.790	0.011	0.358	0.154	0.099	0.997	0.080	0.997	0.193
Panel C: Additional controls											
Treatment	0.004	0.062***	0.021	0.102***	0.037	0.072**	-0.174**	0.014	0.093***	0.028	0.100**
	(0.018)	(0.023)	(0.023)	(0.028)	(0.023)	(0.030)	(0.082)	(0.047)	(0.033)	(0.131)	(0.047)
Observations	2435	2380	2380	2380	2380	2380	2380	2380	2380	2380	2380
Adjusted R-squared	0.330	0.023	0.006	0.021	0.048	0.023	0.300	0.084	0.024	0.008	0.041
Q-statistic	0.866	0.068	0.737	0.009	0.372	0.108	0.128	0.854	0.068	0.866	0.128
Mean control group	0.191	0.000	0.000	-0.002	-0.001	-0.016	-0.606	0.000	0.000	4.513	0.000

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables. Q-statistics are False Discovery Rate corrected q-values based on Benjamini and Hochberg (1995). These are computed pooling all Table 6s results within a panel.

* p<0.10, **p<0.5, *** p<0.01

Note: Marital expectations index is not mean 0 because married girls are assigned the minimum value calculated for non-married girls

Table 6-3: Non-cognitive skills - Parent reports and new indices

	(1) Parental Perception of Girl's Strengths	(2) Parental Perception of Girl's Self-Efficacy	(3) Parental Perception Freedom of Movement	(4) Parent-Daughter Communication	(5) Parental Gender Attitudes	(6) Parental Schooling Attitudes	(7) Parental Marriage Attitudes	(8) Locus of Control Index	(9) Perceived Stress Index	(10) Rosenberg Self-Esteem Index
Panel A: No controls										
Treatment	-0.042**	0.004	0.021	-0.014	0.000	0.032	0.022	-0.015	-0.025	0.016
	(0.018)	(0.029)	(0.029)	(0.029)	(0.026)	(0.042)	(0.031)	(0.046)	(0.047)	(0.030)
Observations	2434	2430	2434	2434	2434	2434	2434	2380	2380	2380
Adjusted R-squared	0.002	-0.001	0.001	0.001	0.009	0.002	0.002	0.000	-0.000	0.005
Q-statistic	0.100	0.956	0.827	0.827	0.991	0.827	0.827	0.925	0.827	0.827
Panel B: Baseline controls										
Treatment	-0.043**	-0.001	0.025	-0.009	0.003	0.027	0.023	-0.020	-0.023	0.024
	(0.018)	(0.030)	(0.028)	(0.028)	(0.026)	(0.038)	(0.031)	(0.046)	(0.047)	(0.030)
Observations	2434	2430	2434	2434	2434	2434	2434	2380	2380	2380
Adjusted R-squared	0.010	0.013	0.006	0.016	0.029	0.105	0.024	0.013	-0.001	0.023
Q-statistic	0.099	0.997	0.790	0.933	0.997	0.790	0.790	0.879	0.872	0.790
Panel C: Additional controls										
Treatment	-0.040**	0.010	0.031	-0.009	0.010	0.043	0.022	-0.026	-0.024	0.021
	(0.018)	(0.029)	(0.028)	(0.029)	(0.026)	(0.037)	(0.031)	(0.046)	(0.047)	(0.030)
Observations	2434	2430	2434	2434	2434	2434	2434	2380	2380	2380
Adjusted R-squared	0.010	0.018	0.006	0.016	0.034	0.117	0.026	0.012	-0.002	0.026
Q-statistic	0.128	0.854	0.629	0.854	0.854	0.617	0.798	0.833	0.833	0.798
Mean control group	0.000	-0.002	0.000	0.000	0.000	0.001	-0.004	0.000	0.000	0.000

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables. Q-statistics are False Discovery Rate corrected q-values based on Benjamini and Hochberg (1995). These are computed pooling all Table 6s results within a panel.
* p<0.10, **p<0.5, *** p<0.01

Table 7: Child labor

	(1) Child Works (Economically Active)	(2) Child Works for Pay	(3) Child Works outside of Family Activity	(4) Child Labor	(5) Hazardous Child Labor	(6) Other Worst Forms of Child Labor	(7) Hours worked in a day	(8) Hours worked in a day - unpaid work	(9) Total Hours Active (Paid + Unpaid)	(10) Hours active outside house
Panel A: No controls										
Treatment	0.049	0.021	-0.011	0.004	0.009	0.021	0.060	0.026	0.086	0.000
	(0.040)	(0.025)	(0.030)	(0.037)	(0.036)	(0.021)	(0.138)	(0.074)	(0.171)	(0.086)
Observations	2386	2386	2387	2386	2386	2387	2386	2386	2386	2386
Adjusted R-squared	0.003	0.008	0.004	0.003	0.002	0.002	0.003	0.001	0.005	0.002
Panel B: Controls for baseline values										
Treatment	0.044	0.023	-0.008	0.006	0.012	0.021	0.001	0.006	-0.005	-0.023
	(0.037)	(0.025)	(0.029)	(0.034)	(0.033)	(0.020)	(0.120)	(0.069)	(0.148)	(0.080)
Observations	2386	2386	2387	2386	2386	2387	2386	2386	2386	2386
Adjusted R-squared	0.037	0.013	0.006	0.026	0.028	0.007	0.121	0.083	0.165	0.059
Panel C: Additional controls										
Treatment	0.031	0.012	-0.008	-0.012	-0.006	0.012	-0.024	-0.009	-0.036	-0.038
	(0.035)	(0.025)	(0.030)	(0.032)	(0.031)	(0.020)	(0.124)	(0.069)	(0.152)	(0.083)
Observations	2386	2386	2387	2386	2386	2387	2386	2386	2386	2386
Adjusted R-squared	0.070	0.024	0.006	0.054	0.053	0.013	0.122	0.086	0.165	0.059
Mean control group	0.651	0.228	0.186	0.583	0.458	0.180	1.157	1.642	2.800	0.602

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables.

* p<0.10, **p<0.5, *** p<0.01

Table 8: Cognitive skills and academic achievement

	(1)	(2)	(3)	(4)	(5)
	Hours studying at home	Total hours spent on school	ASER score - Mathematics	ASER score - Hindi	ASER score - English
Panel A: No controls					
Treatment	-0.062	0.183	-0.021	0.032	-0.074
	(0.077)	(0.189)	(0.077)	(0.093)	(0.090)
Observations	2386	2386	2380	2380	2380
Adjusted R-squared	0.002	0.002	0.002	0.003	0.001
Panel B: Controls for baseline values					
Treatment	-0.067	0.132	-0.032	0.008	-0.089
	(0.076)	(0.187)	(0.070)	(0.089)	(0.084)
Observations	2386	2386	2380	2380	2380
Adjusted R-squared	0.035	0.084	0.062	0.071	0.080
Panel C: Additional controls					
Treatment	-0.043	0.164	-0.014	0.022	-0.068
	(0.075)	(0.188)	(0.070)	(0.089)	(0.084)
Observations	2386	2386	2380	2380	2380
Adjusted R-squared	0.041	0.090	0.065	0.072	0.083
Mean control group	1.541	7.166	2.353	3.025	2.369

Standard errors, clustered by school, in parenthesis. Panel A contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported) and stratification fixed effects. Panel B adds age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified. Panel C adds in controls for variables that appear imbalanced in the balance tables.

* p<0.10, **p<0.5, *** p<0.01

Table 9: School progression and completion, heterogeneous effects

	(1) Whether child has dropped out	(2) Whether child progressed to grade 7th	(3) Attendance rate	(4) Attendance - dummy
School Quality				
Treatment	-0.017	0.019	-0.002	-0.000
	(0.024)	(0.026)	(0.013)	(0.007)
Treatment X School quality	-0.031	0.031	0.013	0.008
	(0.037)	(0.038)	(0.019)	(0.011)
Age				
Treatment	0.019	0.009	-0.053	-0.011
	(0.064)	(0.066)	(0.041)	(0.029)
Treatment X Age	-0.005	0.002	0.005	0.001
	(0.007)	(0.007)	(0.004)	(0.003)
Maternal education				
Treatment	-0.035	0.037	-0.006	0.001
	(0.022)	(0.023)	(0.010)	(0.007)
Treatment X Maternal education	0.009	-0.002	0.057***	0.018
	(0.035)	(0.039)	(0.021)	(0.012)
Economic shock				
Treatment	-0.023	0.037	0.010	-0.000
	(0.024)	(0.023)	(0.016)	(0.011)
Treatment X Economic shock	-0.018	-0.001	-0.008	0.006
	(0.025)	(0.024)	(0.018)	(0.012)
Crime shock				
Treatment	-0.029	0.032	0.014	0.005
	(0.020)	(0.020)	(0.010)	(0.006)
Treatment X Crime shock	-0.026	0.024	-0.069***	-0.009
	(0.034)	(0.035)	(0.025)	(0.014)
Death / illness shock				
Treatment	-0.043*	0.047*	0.005	0.003
	(0.024)	(0.025)	(0.011)	(0.007)
Treatment X Death / illness shock	0.025	-0.027	-0.001	0.003
	(0.026)	(0.028)	(0.017)	(0.010)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 10-1: Non-cognitive skills - Objective measures, heterogeneous effects

	(1) Delay discounting	(2) Completed Mirror Drawings	(3) Mirror Drawings (seconds)	(4) Scavenger Hunt Index
School Quality				
Treatment	0.022	-0.066	-4.313	-0.110
	(0.040)	(0.110)	(6.527)	(0.079)
Treatment X School quality	-0.056	0.232	13.871	0.050
	(0.064)	(0.170)	(9.084)	(0.116)
Age				
Treatment	0.076	0.493	31.321*	0.078
	(0.115)	(0.377)	(18.289)	(0.214)
Treatment X Age	-0.007	-0.040	-2.635	-0.015
	(0.010)	(0.032)	(1.595)	(0.020)
Maternal education				
Treatment	0.002	0.037	3.458	-0.081
	(0.032)	(0.079)	(4.839)	(0.061)
Treatment X Maternal education	-0.032	0.163	-2.609	0.063
	(0.045)	(0.167)	(9.717)	(0.123)
Economic shock				
Treatment	0.040	0.082	1.053	-0.031
	(0.040)	(0.102)	(6.040)	(0.067)
Treatment X Economic shock	-0.069*	-0.035	2.689	-0.082
	(0.038)	(0.093)	(6.758)	(0.079)
Crime shock				
Treatment	-0.001	0.061	4.088	-0.082
	(0.034)	(0.085)	(4.838)	(0.059)
Treatment X Crime shock	-0.006	0.006	-11.037	0.011
	(0.053)	(0.140)	(9.656)	(0.122)
Death / illness shock				
Treatment	0.019	0.078	6.174	-0.054
	(0.033)	(0.093)	(5.875)	(0.063)
Treatment X Death / illness shock	-0.052	-0.038	-8.772	-0.071
	(0.033)	(0.106)	(8.007)	(0.082)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 10-2: Non-cognitive skills - Survey measures, heterogeneous effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Married	Socio-emotional Index	Freedom of Movement Index	Empowerment Index	Self-Esteem Index	Future Planning Index	Marital Expectations Index	Employment Expectations Index	Gender Norms Index	Cantril ladder	Enumerator Assessment Index
School Quality											
Treatment	0.092**	0.058**	-0.023	0.135***	0.066*	0.076	-0.502***	-0.048	0.133***	0.305*	0.114
	(0.042)	(0.028)	(0.032)	(0.040)	(0.034)	(0.046)	(0.175)	(0.072)	(0.051)	(0.157)	(0.077)
Treatment X School quality	-0.097*	0.014	0.074*	-0.071	-0.041	-0.016	0.340	0.056	-0.100	-0.589**	-0.067
	(0.055)	(0.047)	(0.044)	(0.054)	(0.048)	(0.063)	(0.234)	(0.103)	(0.067)	(0.262)	(0.100)
Age											
Treatment	-0.070	-0.128	0.001	0.151	-0.045	0.094	-0.101	-0.048	0.006	-0.603	0.147
	(0.111)	(0.109)	(0.098)	(0.127)	(0.105)	(0.143)	(0.453)	(0.184)	(0.131)	(0.548)	(0.250)
Treatment X Age	0.010	0.018*	0.002	-0.005	0.008	-0.002	-0.020	0.003	0.008	0.054	-0.006
	(0.010)	(0.010)	(0.009)	(0.011)	(0.010)	(0.013)	(0.041)	(0.017)	(0.012)	(0.049)	(0.023)
Maternal education											
Treatment	0.040	0.068***	0.010	0.110***	0.040	0.061*	-0.334***	-0.011	0.095***	-0.104	0.086*
	(0.030)	(0.026)	(0.027)	(0.029)	(0.025)	(0.034)	(0.126)	(0.056)	(0.035)	(0.147)	(0.051)
Treatment X Maternal education	-0.012	-0.001	0.069	-0.072	0.015	0.066	0.142	0.000	-0.017	0.505*	-0.016
	(0.035)	(0.055)	(0.047)	(0.046)	(0.042)	(0.063)	(0.162)	(0.086)	(0.054)	(0.276)	(0.101)
Economic shock											
Treatment	0.045	0.078**	0.032	0.088**	0.004	0.068	-0.344**	0.012	0.126***	0.018	0.079
	(0.034)	(0.031)	(0.035)	(0.035)	(0.034)	(0.045)	(0.146)	(0.062)	(0.038)	(0.174)	(0.066)
Treatment X Economic shock	-0.004	-0.018	-0.021	0.012	0.063	0.003	0.033	-0.048	-0.064	-0.060	-0.002
	(0.033)	(0.033)	(0.040)	(0.037)	(0.043)	(0.052)	(0.140)	(0.064)	(0.045)	(0.177)	(0.080)
Crime shock											
Treatment	0.047	0.078***	0.021	0.091***	0.043*	0.068**	-0.336***	0.006	0.090***	-0.007	0.101*
	(0.028)	(0.026)	(0.024)	(0.029)	(0.026)	(0.034)	(0.119)	(0.055)	(0.034)	(0.129)	(0.053)
Treatment X Crime shock	-0.040	-0.077	-0.001	0.030	-0.012	0.012	0.129	-0.168*	-0.005	-0.055	-0.163*
	(0.049)	(0.058)	(0.047)	(0.058)	(0.051)	(0.065)	(0.217)	(0.085)	(0.067)	(0.295)	(0.092)

Death / illness shock											
Treatment	-0.040	-0.077	-0.001	0.030	-0.012	0.012	0.129	-0.168*	-0.005	-0.055	-0.163*
	(0.049)	(0.058)	(0.047)	(0.058)	(0.051)	(0.065)	(0.217)	(0.085)	(0.067)	(0.295)	(0.092)
Treatment X Death / illness shock	0.054*	0.001	-0.003	-0.050	-0.017	-0.059	0.010	-0.002	0.023	-0.120	0.004
	(0.031)	(0.039)	(0.039)	(0.034)	(0.035)	(0.049)	(0.122)	(0.060)	(0.045)	(0.185)	(0.066)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 10-3: Non-cognitive skills - Parent reports and new indices, heterogeneous effects

	(1) Parental Perception of Girl's Strengths	(2) Parental Perception of Girl's Self-Efficacy	(3) Parental Perception Freedom of Movement	(4) Parent-Daughter Communication	(5) Parental Gender Attitudes	(6) Parental Schooling Attitudes	(7) Parental Marriage Attitudes	(8) Locus of Control Index	(9) Perceived Stress Index	(10) Rosenberg Self-Esteem Index
School Quality										
Treatment	-0.013	0.029	0.038	0.057	0.039	0.074	0.066	-0.018	-0.038	0.064
	(0.024)	(0.044)	(0.037)	(0.042)	(0.037)	(0.051)	(0.044)	(0.070)	(0.065)	(0.045)
Treatment X School quality	-0.058	-0.052	-0.042	-0.134**	-0.074	-0.093	-0.085	0.014	0.027	-0.087
	(0.037)	(0.060)	(0.056)	(0.056)	(0.052)	(0.081)	(0.061)	(0.094)	(0.094)	(0.060)
Age										
Treatment	-0.062	-0.076	0.295**	-0.011	0.099	0.159	-0.136	0.001	-0.017	-0.016
	(0.083)	(0.146)	(0.125)	(0.113)	(0.091)	(0.155)	(0.133)	(0.211)	(0.236)	(0.112)
Treatment X Age	0.002	0.007	-0.025**	0.000	-0.009	-0.012	0.014	-0.002	-0.001	0.003
	(0.007)	(0.013)	(0.011)	(0.010)	(0.008)	(0.014)	(0.012)	(0.019)	(0.021)	(0.010)
Maternal education										
Treatment	-0.048**	0.009	0.018	-0.020	0.015	0.044	0.025	-0.002	0.000	0.014
	(0.020)	(0.033)	(0.031)	(0.029)	(0.027)	(0.042)	(0.034)	(0.049)	(0.049)	(0.030)
Treatment X Maternal education	0.048	0.020	0.041	0.077	-0.043	-0.023	0.001	-0.124	-0.205*	0.028
	(0.044)	(0.062)	(0.049)	(0.047)	(0.053)	(0.072)	(0.066)	(0.104)	(0.115)	(0.058)
Economic shock										
Treatment	-0.055*	-0.020	0.034	-0.011	-0.001	0.074	0.061	-0.013	-0.044	0.009
	(0.028)	(0.040)	(0.040)	(0.037)	(0.033)	(0.054)	(0.043)	(0.064)	(0.079)	(0.040)
Treatment X Economic shock	0.021	0.043	-0.019	-0.001	0.005	-0.071	-0.066	-0.009	0.022	0.017
	(0.030)	(0.050)	(0.045)	(0.036)	(0.034)	(0.054)	(0.049)	(0.089)	(0.098)	(0.042)
Crime shock										
Treatment	-0.033	0.006	0.017	-0.007	0.009	0.021	0.022	-0.038	-0.040	0.021
	(0.020)	(0.031)	(0.031)	(0.029)	(0.028)	(0.044)	(0.033)	(0.048)	(0.051)	(0.030)
Treatment X Crime shock	-0.072	0.001	0.047	-0.028	-0.050	0.079	-0.000	0.175	0.083	-0.011
	(0.045)	(0.077)	(0.060)	(0.051)	(0.046)	(0.082)	(0.063)	(0.127)	(0.124)	(0.054)

Death / illness shock										
Treatment	-0.023	0.022	0.005	-0.007	-0.006	0.030	0.029	-0.032	-0.026	0.015
	(0.022)	(0.037)	(0.033)	(0.031)	(0.031)	(0.050)	(0.038)	(0.055)	(0.055)	(0.035)
Treatment X Death / illness shock	-0.048*	-0.040	0.046	-0.011	0.018	0.001	-0.018	0.041	-0.008	0.008
	(0.028)	(0.048)	(0.042)	(0.037)	(0.036)	(0.059)	(0.042)	(0.080)	(0.087)	(0.038)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 11: Child labor, heterogeneous effects

	(1) Child Works (Economically Active)	(2) Child Works for Pay	(3) Child Works outside of Family Activity	(4) Child Labor	(5) Hazardous Child Labor	(6) Other Worst Forms of Child Labor	(7) Hours worked in a day	(8) Hours worked in a day - unpaid work	(9) Total Hours Active (Paid + Unpaid)	(10) Hours active outside house
School Quality										
Treatment	0.061	0.040	0.028	0.020	0.016	-0.010	0.162	-0.075	0.086	0.058
	(0.055)	(0.031)	(0.041)	(0.050)	(0.044)	(0.029)	(0.179)	(0.111)	(0.226)	(0.108)
Treatment X School quality	-0.010	-0.035	-0.068	-0.018	0.003	0.059	-0.149	0.213	0.064	-0.088
	(0.076)	(0.050)	(0.060)	(0.070)	(0.067)	(0.041)	(0.265)	(0.144)	(0.328)	(0.168)
Age										
Treatment	0.036	0.136	-0.025	-0.008	0.015	0.024	0.123	0.086	0.210	0.428
	(0.120)	(0.123)	(0.111)	(0.130)	(0.105)	(0.093)	(0.507)	(0.326)	(0.545)	(0.345)
Treatment X Age	0.002	-0.010	0.001	0.001	-0.000	-0.000	-0.004	-0.006	-0.011	-0.038
	(0.011)	(0.011)	(0.010)	(0.012)	(0.009)	(0.009)	(0.054)	(0.031)	(0.057)	(0.036)
Maternal education										
Treatment	0.041	0.011	-0.025	-0.007	-0.001	0.025	0.070	0.018	0.088	-0.003
	(0.037)	(0.028)	(0.030)	(0.035)	(0.035)	(0.022)	(0.154)	(0.078)	(0.188)	(0.095)
Treatment X Maternal education	0.045	0.063	0.094*	0.074	0.074	-0.025	-0.066	-0.040	-0.106	0.021
	(0.057)	(0.052)	(0.052)	(0.056)	(0.053)	(0.039)	(0.199)	(0.147)	(0.269)	(0.131)
Economic shock										
Treatment	0.034	0.023	-0.034	-0.022	-0.016	0.015	0.073	-0.023	0.050	0.013
	(0.042)	(0.030)	(0.035)	(0.042)	(0.042)	(0.031)	(0.163)	(0.103)	(0.204)	(0.096)
Treatment X Economic shock	0.031	-0.003	0.041	0.049	0.050	0.011	-0.003	0.068	0.066	-0.013
	(0.040)	(0.034)	(0.033)	(0.042)	(0.042)	(0.037)	(0.160)	(0.117)	(0.197)	(0.122)
Crime shock										
Treatment	0.052	0.019	-0.015	0.004	0.012	0.015	0.090	0.003	0.093	0.015
	(0.038)	(0.027)	(0.030)	(0.036)	(0.035)	(0.022)	(0.142)	(0.076)	(0.178)	(0.091)
Treatment X Crime shock	0.000	0.028	0.045	0.025	0.012	0.045	-0.142	0.108	-0.034	-0.043
	(0.055)	(0.050)	(0.043)	(0.057)	(0.062)	(0.054)	(0.257)	(0.164)	(0.313)	(0.152)

Death / illness shock										
Treatment	0.040	0.009	-0.040	-0.007	-0.003	0.031	0.023	-0.025	-0.002	0.001
	(0.038)	(0.029)	(0.032)	(0.035)	(0.035)	(0.022)	(0.156)	(0.089)	(0.197)	(0.106)
Treatment X Death / illness shock	0.029	0.034	0.075**	0.033	0.042	-0.025	0.115	0.103	0.218	0.017
	(0.039)	(0.031)	(0.035)	(0.041)	(0.041)	(0.030)	(0.168)	(0.122)	(0.222)	(0.125)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 12: Cognitive skills and academic achievement, heterogeneous effects

	(1) Hours studying at home	(2) Total hours spent on school	(3) ASER score - Mathematics	(4) ASER score - Hindi	(5) ASER score - English
School Quality					
Treatment	-0.141	-0.118	-0.020	-0.095	-0.028
	(0.104)	(0.223)	(0.121)	(0.147)	(0.136)
Treatment X School quality	0.135	0.510	0.004	0.215	-0.112
	(0.156)	(0.363)	(0.158)	(0.187)	(0.182)
Age					
Treatment	0.212	0.259	-0.124	0.248	-0.038
	(0.312)	(0.647)	(0.275)	(0.322)	(0.282)
Treatment X Age	-0.026	-0.008	0.010	-0.020	-0.004
	(0.028)	(0.062)	(0.024)	(0.029)	(0.025)
Maternal education					
Treatment	-0.050	0.153	-0.032	0.040	-0.093
	(0.087)	(0.209)	(0.079)	(0.103)	(0.094)
Treatment X Maternal education	-0.075	0.151	0.129	0.025	0.158
	(0.158)	(0.327)	(0.120)	(0.137)	(0.126)
Economic shock					
Treatment	-0.025	0.252	-0.091	-0.007	-0.100
	(0.101)	(0.217)	(0.091)	(0.108)	(0.106)
Treatment X Economic shock	-0.077	-0.139	0.124	0.061	0.038
	(0.116)	(0.255)	(0.091)	(0.121)	(0.112)
Crime shock					
Treatment	-0.073	0.156	-0.002	0.057	-0.029
	(0.079)	(0.189)	(0.081)	(0.098)	(0.094)
Treatment X Crime shock	0.029	0.104	-0.128	-0.211	-0.373**
	(0.142)	(0.300)	(0.114)	(0.148)	(0.144)
Death / illness shock					
Treatment	-0.015	0.280	-0.027	0.098	-0.030
	(0.088)	(0.231)	(0.083)	(0.104)	(0.092)
Treatment X Death / illness shock	-0.137	-0.272	0.021	-0.168	-0.120
	(0.113)	(0.251)	(0.079)	(0.110)	(0.092)

Standard errors, clustered by school, in parenthesis. Table contains results from regressing the outcome variable indicated by the column header on an indicator for treatment (reported), the interaction of treatment with the specified characteristics (reported), the characteristics, stratification fixed effects, age fixed effects, baseline value of the outcome, and a vector of dummies for the most important type of employment in the household at baseline. This specification was prespecified.

* p<0.10, **p<0.5, *** p<0.01

Table 13: Life Skills Details

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*	*Whrere taught*
Explicit, significant, large effect					
Can talk to parents about marriage	0.303 (0.017)	0.104 (0.028)	*	34.5	Session 4: communication; session 30: importance of
Has a goal for next week	0.285 (0.017)	0.060 (0.025)	*	21.2	Session 10: goals
Articulated step(s) to reach goal for next week	0.283 (0.017)	0.062 (0.024)	*	21.9	Session 10: goals
Marriage keeps girls from getting schooling	0.388 (0.019)	0.079 (0.024)	*	20.3	Session 3: values; session 19: gender; session 30: education
Safety concerns keep girls from getting schooling	0.063 (0.008)	0.040 (0.013)	*	63.3	Session 3: values; session 19: gender; session 30: education
Distance from school keep girls from getting schooling	0.155 (0.010)	0.055 (0.018)	*	35.2	Session 3: values; session 19: gender; session 30: education
Lack of transport keep girls from getting schooling	0.018 (0.004)	0.013 (0.006)	*	75.2	Session 3: values; session 19: gender; session 30: education
Parents' lack of interest keeps girls from getting schooling	0.316 (0.015)	0.054 (0.023)	*	17.1	Session 3: values; session 19: gender; session 30: education
There are no issues keeping girls from getting schooling	0.112 (0.014)	-0.053 (0.016)	*	-47.1	Session 3: values; session 19: gender; session 30: education
Non-conducive social environment keeps girls from getting schooling	0.025 (0.005)	0.022 (0.008)	*	87.1	Session 3: values; session 19: gender; session 30: education
Inlaws' lack of interest keeps girls from getting schooling	0.035 (0.007)	0.026 (0.011)	*	75.3	Session 3: values; session 19: gender; session 30: education
Explicit, significant, small effect					
Has a place to meet female friends	0.470 (0.018)	0.065 (0.028)	*	13.7	Session 24: safe spaces for girls
Has someone to talk about a problem	0.839 (0.010)	0.037 (0.015)	*	4.4	Session 12: friendship; Session 6: feelings
Can talk to parents about future work	0.801 (0.014)	0.057 (0.018)	*	7.1	Session 31: careers; Session 9: decision making; session 5:
Household chores keep girls from getting schooling	0.664 (0.015)	0.063 (0.022)	*	9.4	Session 3: values; session 19: gender; session 30: education
Boys and girls should do the same amount of household chores	0.891 (0.010)	0.033 (0.012)	*	3.7	Session 19: gender; session 3: values
Girls shoud have the same freedom as boys	0.849 (0.015)	0.050 (0.019)	*	5.9	Session 19: gender; session 3: values
When sad, can find something to do that helps	0.851 (0.011)	0.053 (0.014)	*	6.2	Session 6: identifying and expressing feelings
Considerate of other people's feelings ("Certainly true")	0.692 (0.016)	-0.053 (0.021)	*	-7.6	Session 22: empathy
Thinks before acting ("Certainly true")	0.628 (0.019)	-0.082 (0.025)	*	-13.1	Session 9: decision making; session 7: how to manage anger
Even if girls will get married, they should be sent for higher education	0.781 (0.016)	0.045 (0.022)	*	5.8	Session 19: gender; session 3: values; session 30: education

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*	*% Change*	*Whrere taught*
Explicit, insignificant, large effect				
Has an education-related goal for next week	0.176 (0.013)	0.034 (0.018)	19.4	Session 10: goals
Least amount of expected education is tertiary education	0.316 (0.021)	-0.051 (0.028)	-16.1	Session 30: education; session 10: goals
Relocation keep girls from getting schooling	0.014 (0.003)	-0.006 (0.005)	-45.8	Session 3: values; session 19: gender; session 30: education
Household work keep girls from getting schooling	0.098 (0.008)	0.015 (0.013)	15.6	Session 3: values; session 19: gender; session 30: education
Girls' involvement in unsocial practices keeps girls from schooling	0.020 (0.004)	0.011 (0.007)	55.3	Session 3: values; session 19: gender; session 30: education
Family death keeps girls from getting schooling	0.012 (0.003)	-0.004 (0.004)	-34.3	Session 3: values; session 19: gender; session 30: education
Vocational training keep girls from getting schooling	0.001 (0.001)	-0.001 (0.001)	-100.3	Session 3: values; session 19: gender; session 30: education
Girls' illness keeps girls from getting schooling	0.003 (0.002)	0.003 (0.003)	78.8	Session 3: values; session 19: gender; session 30: education
Being admitted to other school keeps girls from getting schooling	0.003 (0.001)	-0.001 (0.002)	-30.8	Session 3: values; session 19: gender; session 30: education
Fear of teachers keeps girls from getting schooling	0.004 (0.002)	-0.003 (0.002)	-64.8	Session 3: values; session 19: gender; session 30: education
Social pressures keep girls from getting schooling	0.039 (0.006)	0.009 (0.009)	24.1	Session 3: values; session 19: gender; session 30: education
Family migration keeps girls from getting schooling	0.008 (0.003)	0.005 (0.005)	63.1	Session 3: values; session 19: gender; session 30: education
Parents' death keeps girls from getting schooling	0.011 (0.003)	0.010 (0.005)	96.3	Session 3: values; session 19: gender; session 30: education
Explicit insignificant, small effect				
Has three or more good friends	0.463 (0.018)	0.023 (0.028)	5.0	Session 11: teambuilding; session 12: friendship
Met friend outside school in last week	0.635 (0.015)	0.042 (0.023)	6.7	Session 12: friendship
Usually/always has peer who will listen if she needs to talk	0.553 (0.017)	0.018 (0.024)	3.3	Session 11: teambuilding; session 12: friendship
Usually/always has someone to share accomplishments with	0.491 (0.017)	0.006 (0.023)	1.3	Session 11: teambuilding; session 12: friendship
Usually/always has a friend ask her for help/advice	0.339 (0.015)	0.022 (0.024)	6.4	Session 12: friendship; session 13: helping a friend
Has gone to school in the past month	0.884 (0.017)	0.033 (0.019)	3.8	Session 30: importance of education
Can talk to parents about how much schooling she'll have	0.890 (0.011)	0.029 (0.015)	3.3	Session 30: importance of education; Session 9: decision
Feels as important as other family members	0.954 (0.008)	-0.010 (0.011)	-1.0	Session 2: self awareness; session 3: values
Likes to make pans for her future work/studies	0.861 (0.012)	0.024 (0.017)	2.8	Session 31: careers; session 10: goals
Hopeful about her future	0.905 (0.010)	0.012 (0.013)	1.3	Session 10: goals; session 11: reaching goals together
Can solve most problems if she tries hard enough	0.877 (0.012)	0.001 (0.016)	0.1	Session 8: problem solving
Wants to become like a professional when she grows up	0.691 (0.018)	0.038 (0.024)	5.5	Session 31: careers; session 10: goals

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*	*% Change*	*Whrere taught*
Has an education-related goal for the next year	0.258 (0.013)	0.032 (0.019)	12.4	Session 10: goals
Has a goal for next yar	0.540 (0.018)	0.028 (0.024)	5.2	Session 10: goals
Articulated step(s) to reach goal for next year	0.541 (0.018)	0.027 (0.024)	5.0	Session 10: goals
Has thought about goals	0.853 (0.014)	0.029 (0.017)	3.4	Session 10: goals
When she's excited about reaching a goal, it's easy to start	0.698 (0.019)	0.033 (0.025)	4.8	Session 10: goals
Can usually find a way to stick with plans/goals	0.891 (0.013)	-0.010 (0.019)	-1.1	Session 10: goals; session 11: reaching goals together
Hopes to have a job for pay in the future	0.759 (0.018)	0.013 (0.023)	1.8	Session 31: careers; session 10: goals
Hopes to have a professional job in the future	0.685 (0.020)	-0.005 (0.026)	-0.7	Session 31: careers; session 10: goals
Least amount of expected education is more than no further schooling	0.947 (0.008)	0.002 (0.011)	0.2	Session 30: education; session 10: goals
Least amount of expected education is secondary school or more	0.922 (0.012)	0.008 (0.015)	0.9	Session 30: education; session 10: goals
Greatest amount of expected education is more than no further schooling	0.947 (0.008)	0.002 (0.011)	0.2	Session 30: education; session 10: goals
Greatest amount of expected education is secondary school or more	0.935 (0.010)	0.000 (0.013)	0.0	Session 30: education; session 10: goals
Greatest amount of expected education is tertiary education	0.567 (0.025)	0.005 (0.031)	0.9	Session 30: education; session 10: goals
After current school, will study at a different school	0.639 (0.024)	0.053 (0.033)	8.3	Session 31: careers; session 10: goals
Lack of money keeps girls from getting schooling	0.554 (0.017)	0.015 (0.023)	2.6	Session 3: values; session 19: gender; session 30: education
Need to work on farm keeps girls from getting schooling	0.287 (0.022)	0.036 (0.032)	12.7	Session 3: values; session 19: gender; session 30: education
Lack of interest keeps girls from getting schooling	0.199 (0.013)	0.001 (0.017)	0.6	Session 3: values; session 19: gender; session 30: education
Child not being good at studies keeps girls from getting schooling	0.104 (0.010)	0.011 (0.014)	10.6	Session 3: values; session 19: gender; session 30: education
Girls' illness keeps girls from getting schooling	0.017 (0.004)	0.002 (0.006)	12.4	Session 3: values; session 19: gender; session 30: education
Boys should not be sent to school before girls if money is scare	0.770 (0.018)	0.019 (0.024)	2.5	Session 19: gender; session 3: values; session 30: education
Boys are not naturally better at studying than girls	0.786 (0.018)	0.042 (0.023)	5.4	Session 19: gender; session 3: values; session 30: education
Helpful if someone is hurt/upset/ill ("Certainly true")	0.753 (0.014)	-0.030 (0.019)	-4.0	Session 13: how to help a friend; session 22: empathy
Has at least one good friend ("Certainly true")	0.819 (0.015)	-0.029 (0.020)	-3.5	Session 12: friendship
Kind to younger children ("Certainly true")	0.868 (0.010)	-0.027 (0.016)	-3.1	Session 22: empathy
Does not often lose temper	0.327 (0.016)	0.005 (0.022)	1.6	Session 2: self-awareness; session 7: how to manage anger
Sets goal and follows through necessary steps to achieve (sometimes-almost always)	0.820 (0.012)	0.020 (0.017)	2.4	Session 10: goals; session 8: problem solving
Remains calm when facing difficulties (sometimes-almost always)	0.704 (0.020)	-0.002 (0.027)	-0.2	Session 2: self-awareness; session 7: how to manage anger

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*	*Whrere taught*
Makes her opinions known about things that affect her (sometimes-almost always)	0.852 (0.013)	0.012 (0.017)		1.4	Session 4: communication; session 5: assertiveness
Discussed something that was worrying her with parents	0.551 (0.022)	-0.031 (0.031)		-5.6	Session 6: identifying and expressing feelings; Session 4:
Discussed something that was going well with parents	0.718 (0.018)	0.002 (0.026)		0.3	Session 6: identifying and expressing feelings; Session 4:
Discussed plans and goals for education with parents	0.692 (0.018)	-0.020 (0.024)		-2.9	Session 10: goals; session 30: importance of education
Discussed future work with parents	0.475 (0.017)	0.010 (0.024)		2.0	Session 31: careers
Indirect, significant, large effect					
Went to health center alone at least sometimes	0.077 (0.013)	-0.043 (0.016)	*	-55.5	Not explicitly in the lesson plans
Went to fest alone at least sometimes	0.043 (0.010)	0.044 (0.017)	*	101.7	Not explicitly in the lesson plans
Allowed to go to mela alone at least sometimes	0.038 (0.006)	0.020 (0.009)	*	52.6	Not explicitly in the lesson plans
Allowed to go to school alone at least sometimes	0.035 (0.005)	0.025 (0.008)	*	71.9	Not explicitly in the lesson plans
Child mostly decides whether she goes to school	0.412 (0.015)	0.068 (0.024)	*	16.6	Not explicitly in the lesson plans
Child mostly decides whether she will continue school after 8th grade	0.328 (0.016)	0.054 (0.022)	*	16.6	Not explicitly in the lesson plans
Child mostly decides when she'll move in with in-laws	0.035 (0.005)	0.025 (0.010)	*	70.1	Not explicitly in the lesson plans
Child mostly decides if she will work after her studies	0.331 (0.016)	0.064 (0.023)	*	19.3	Not explicitly in the lesson plans
Child mostly decides the type of work she'll do after studies	0.375 (0.017)	0.099 (0.025)	*	26.5	Not explicitly in the lesson plans
Able to talk to parents about preferences for who she marries	0.270 (0.015)	0.080 (0.023)	*	29.7	Not explicitly in the lesson plans
Allowed to go to a mela alone at least sometimes	0.015 (0.004)	0.015 (0.006)	*	99.6	Not explicitly in the lesson plans

Curricular means the question is directly addressed in the Level 6 or Level 7 Curriculum. A large effect is at least a 15 percent change. * indicates q-values less than 0.25 and p-value less than 0.05. is defined as "significant". Q-statistics are False Discovery Rate corrected q-values based on Benjamini and Hochberg (1995). These are computed pooling 259 life skills questions.

Table 14: Life Skills Targets

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*
Self confidence				
Index	0.001	0.028		
	(0.019)	(0.024)		
Feels as important as other family members	0.954	-0.010		-1.0
	(0.008)	(0.011)		
Feels good at math	0.660	0.034		5.1
	(0.018)	(0.025)		
Feels good at reading	0.853	0.036	*	4.3
	(0.012)	(0.016)		
Is comfortable when her teacher calls on her to answer questions	0.826	-0.002		-0.3
	(0.014)	(0.018)		
Hopeful about her future	0.905	0.012		1.3
	(0.010)	(0.013)		
Can solve most problems if she tries hard enough	0.877	0.001		0.1
	(0.012)	(0.016)		
Girls shoud have the same freedom as boys	0.849	0.050	*	5.9
	(0.015)	(0.019)		
Is not nervous in new situation/does not lose confidence easily	0.366	-0.018		-5.0
	(0.017)	(0.023)		
Eager to learn new things (sometimes-almost always)	0.891	0.006		0.6
	(0.008)	(0.012)		
Expressing and managing emotion				
Index	0.001	0.018		
	(0.014)	(0.021)		
Has someone to talk about a problem	0.839	0.037	*	4.4
	(0.010)	(0.015)		
When sad, can find something to do that helps	0.851	0.053	*	6.2
	(0.011)	(0.014)		
Does not have trouble forcing herself to pay attention in a dull class	0.323	-0.034		-10.5
	(0.015)	(0.021)		
Does not have many worries/seem worried	0.393	-0.037		-9.5
	(0.016)	(0.021)		
Is not often unhappy/depressed/tearful	0.570	-0.009		-1.7
	(0.017)	(0.024)		
Empathy				
Index	-0.002	-0.044		
	(0.018)	(0.027)		
Usually/always has a friend ask her for help/advice	0.339	0.022		6.4
	(0.015)	(0.024)		
Shares readily with other youth ("Certainly true")	0.638	-0.011		-1.7
	(0.017)	(0.026)		
Helpful if someone is hurt/upset/ill ("Certainly true")	0.753	-0.030		-4.0
	(0.014)	(0.019)		
Kind to younger children ("Certainly true")	0.868	-0.027		-3.1
	(0.010)	(0.016)		
Often offers to help others ("Certainly true")	0.813	-0.040		-4.9
	(0.014)	(0.020)		

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*
Self-control				
Index	-0.001	-0.064	*	
	(0.015)	(0.021)		
Prefers 60 ruppees a week from now to 30 today	0.331	-0.000		-0.1
	(0.021)	(0.032)		
Thinks before acting ("Certainly true")	0.628	-0.082	*	-13.1
	(0.019)	(0.025)		
Good attention span ("Certainly true")	0.708	-0.029		-4.2
	(0.015)	(0.020)		
Not restless or overactive	0.505	-0.013		-2.5
	(0.016)	(0.025)		
Is not easily distracted	0.386	-0.033		-8.5
	(0.016)	(0.021)		
Remains calm when facing difficulties (sometimes-almost always)	0.704	-0.002		-0.2
	(0.020)	(0.027)		
Kind to younger children ("Certainly true")	0.868	-0.027		-3.1
	(0.010)	(0.016)		
Often offers to help others ("Certainly true")	0.813	-0.040		-4.9
	(0.014)	(0.020)		
Does not have trouble forcing herself to pay attention in a dull class	0.323	-0.034		-10.5
	(0.015)	(0.021)		
Decision making				
Index	-0.012	0.077	*	
	(0.017)	(0.024)		
Child mostly decides whether she can socialize outside house	0.345	0.004		1.2
	(0.017)	(0.024)		
Child mostly decides or jointly decides whether she can socialize outside house	0.658	0.000		0.1
	(0.023)	(0.032)		
Child mostly decides whether she goes to school	0.412	0.068	*	16.6
	(0.015)	(0.024)		
Child mostly decides or jointly decides whether she goes to school	0.674	0.035		5.2
	(0.022)	(0.029)		
Child mostly decides whether she will continue school after 8th grade	0.328	0.054	*	16.6
	(0.016)	(0.022)		
Child mostly decides or jointly decides whether she will continue school after 8th grade	0.652	0.031		4.8
	(0.025)	(0.030)		
Child mostly decides who she will marry	0.030	0.013		42.0
	(0.005)	(0.009)		
Child mostly decides or jointly decides who she will marry	0.427	0.003		0.6
	(0.023)	(0.031)		
Child mostly decides when she'll get married	0.038	0.018		47.2
	(0.006)	(0.009)		
Child mostly decides or jointly decides when she'll get married	0.430	0.019		4.5
	(0.024)	(0.032)		
Child mostly decides when she'll move in with in-laws	0.035	0.025	*	70.1
	(0.005)	(0.010)		
Child mostly decides or jointly decides when she'll move in with in-laws	0.434	0.026		5.9
	(0.023)	(0.031)		
Child mostly decides if she will work after her studies	0.331	0.064	*	19.3
	(0.016)	(0.023)		
Child mostly decides or jointly decides if she will work after her studies	0.651	0.040		6.2
	(0.021)	(0.027)		
Child mostly decides the type of work she'll do after studies	0.375	0.099	*	26.5
	(0.017)	(0.025)		

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*
Child mostly decides or jointly decides what type of work she'll do after studies	0.681 (0.021)	0.058 (0.027)	*	8.5
Child mostly decides how she spends her free time	0.664 (0.019)	0.074 (0.024)	*	11.2
Child mostly decides or jointly decides how she spends her free time	0.820 (0.017)	0.046 (0.020)	*	5.6
Child mostly decides types of chores she does at home	0.665 (0.020)	0.012 (0.027)		1.8
Child mostly decides or jointly decides types of chores she does at home	0.807 (0.017)	0.028 (0.022)		3.5
Child mostly decides how often she spends time with friends	0.613 (0.018)	0.040 (0.025)		6.5
Child mostly decides or jointly decides how often she spends time with friends	0.798 (0.016)	0.034 (0.022)		4.2
Likes to make pans for her future work/studies	0.861 (0.012)	0.024 (0.017)		2.8
Has a goal for next week	0.285 (0.017)	0.060 (0.025)	*	21.2
Has a goal for next yar	0.540 (0.018)	0.028 (0.024)		5.2
Articulated step(s) to reach goal for next week	0.283 (0.017)	0.062 (0.024)	*	21.9
Articulated step(s) to reach goal for next year	0.541 (0.018)	0.027 (0.024)		5.0
Has thought about goals	0.853 (0.014)	0.029 (0.017)		3.4
Can usually find a way to stick with plans/goals	0.891 (0.013)	-0.010 (0.019)		-1.1
Hopes to have a job for pay in the future	0.759 (0.018)	0.013 (0.023)		1.8
Hopes to have a professional job in the future	0.685 (0.020)	-0.005 (0.026)		-0.7
Sets goal and follows through necessary steps to achieve (sometimes-almost always)	0.820 (0.012)	0.020 (0.017)		2.4
Hopeful about her future	0.905 (0.010)	0.012 (0.013)		1.3
Thinks before acting ("Certainly true")	0.628 (0.019)	-0.082 (0.025)	*	-13.1
Perseverance				
Index	0.001 (0.021)	0.026 (0.033)		
Tries to find a different way to say something if someone doesn't understand	0.836 (0.010)	0.045 (0.015)	*	5.3
Spends extra time and effort to get something hard right	0.881 (0.010)	0.012 (0.016)		1.4
When she succeeds at something, it is because she worked hard	0.884 (0.010)	-0.001 (0.014)		-0.1
Will try to find a way to get someone who opposes her to see her point of view (sometimes-almost always)	0.622 (0.013)	0.019 (0.023)		3.0
Can solve most problems if she tries hard enough	0.877 (0.012)	0.001 (0.016)		0.1
Can usually find a way to stick with plans/goals	0.891 (0.013)	-0.010 (0.019)		-1.1

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*
Communication				
Index	-0.003	0.044		
	(0.018)	(0.025)		
Can talk to parents about how much schooling she'll have	0.890	0.029		3.3
	(0.011)	(0.015)		
Able to talk to parents about preferences for who she marries	0.270	0.080	*	29.7
	(0.015)	(0.023)		
Can talk to parents about marriage	0.303	0.104	*	34.5
	(0.017)	(0.028)		
Can talk to parents about future work	0.801	0.057	*	7.1
	(0.014)	(0.018)		
Can talk to parents when having problems with friends or at school	0.849	0.028		3.3
	(0.011)	(0.015)		
Makes her opinions known about things that affect her (sometimes-almost always)	0.852	0.012		1.4
	(0.013)	(0.017)		
Discussed something that was worrying her with parents	0.551	-0.031		-5.6
	(0.022)	(0.031)		
Discussed something that was going well with parents	0.718	0.002		0.3
	(0.018)	(0.026)		
Discussed plans and goals for education with parents	0.692	-0.020		-2.9
	(0.018)	(0.024)		
Discussed preferences for when she'll be married with parents	0.059	0.004		6.3
	(0.009)	(0.012)		
Discussed preferences about who she'll marry with parents	0.039	0.000		0.1
	(0.006)	(0.008)		
Discussed future work with parents	0.475	0.010		2.0
	(0.017)	(0.024)		
Talked to parents about a fight with peers/problem at school	0.531	0.012		2.3
	(0.018)	(0.026)		
Is comfortable when her teacher calls on her to answer questions	0.826	-0.002		-0.3
	(0.014)	(0.018)		
Relationship building				
Index	-0.010	0.050	*	
	(0.013)	(0.020)		
Has three or more good friends	0.463	0.023		5.0
	(0.018)	(0.028)		
Met friend outside school in last week	0.635	0.042		6.7
	(0.015)	(0.023)		
Has a place to meet female friends	0.470	0.065	*	13.7
	(0.018)	(0.028)		
Has someone to take her in if she needs a place to stay for a night	0.759	0.034		4.4
	(0.013)	(0.017)		
Has someone to borrow money form in an emergency	0.790	0.025		3.2
	(0.010)	(0.017)		
Usually/always has peer who will listen if she needs to talk	0.553	0.018		3.3
	(0.017)	(0.024)		
Usually/always has someone to share accomplishments with	0.491	0.006		1.3
	(0.017)	(0.023)		
Has gone to a friend's house in the past month	0.690	0.027		3.9
	(0.016)	(0.025)		
Has at least one good friend ("Certainly true")	0.819	-0.029		-3.5
	(0.015)	(0.020)		
Would not rather be alone than with other youth	0.408	-0.009		-2.3
	(0.016)	(0.025)		

Variable description	*Control Mean / (SE)*	*Treatment Mean / (SE)*		*% Change*
Has someone to talk about a problem	0.839	0.037	*	4.4
	(0.010)	(0.015)		
Usually/always has a friend ask her for help/advice	0.339	0.022		6.4
	(0.015)	(0.024)		
Creative problem solving				
Index	-0.002	0.049		
	(0.018)	(0.030)		
Tries to find a different way to say something if someone doesn't understand	0.836	0.045	*	5.3
	(0.010)	(0.015)		
Spends extra time and effort to get something hard right	0.881	0.012		1.4
	(0.010)	(0.016)		
Will try to find a way to get someone who opposes her to see her point of view (sometimes-almost always)	0.622	0.019		3.0
	(0.013)	(0.023)		
Remains calm when facing difficulties (sometimes-almost always)	0.704	-0.002		-0.2
	(0.020)	(0.027)		

* indicates q-values less than 0.25 and p-value less than 0.05. Q-statistics are False Discovery Rate corrected q-values based on Benjamini and Hochberg (1995). These are computed pooling the life skills questions and indexes displayed.

Table 15: Summary statistics for sampled households

Variable	Sample mean	Rajasthan mean	India mean
Enrollment: girls 10-12	97.5***	70.46	85.22
Enrollment: boys 10-12	97.7***	89.76	89.21
Enrollment: girls 12-14	91.98***	60.69	79.5
Enrollment: boys 12-14	92.37***	82.87	83.71
Child works: girls 10-12	85.6***	90.15	74.03
Land owned (hectare)	1.04***	1.55	2.24
Household size	6.76***	5.41	4.9
Percent Muslim	21.42***	9.85	12.25
Percent Schedule Caste/ Scheduled Tribe	25***	32	31.32
Percent OBC	68***	45	32.72

Notes: *** indicates the difference between the sample mean and state mean is significant at the one percent level

Source: Table from baseline report. Baseline data and the 2007-2008 District Level health Facility Survey (Rajasthan and all-India columns)