Analysis Plan and Variable Definitions

Section A: Pre-Analysis Plan (February 2016)

This trial was registered with the American Economic Association, and the pre-analysis plan included in this Online Appendix was registered prior to baseline (in February 2016). The pre-analysis plan specifies variable construction and our empirical approach, including balance check specifications and models for estimating average treatment effects as well as heterogeneous treatment effects.

Sample Description

The sample includes 119 schools that are located in the Ajmer District of Rajasthan, India. The sample consists of those schools in Ajmer that had between 16 and 32 girls enrolled in class 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 full analysis sample will include all female students who were currently enrolled in class five in these schools as of January 2016 (2,551 female students in total). We conducted a stratified randomization that assigned 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. All girls in Treatment schools will receive the Room to Read intervention during the school year beginning in June, 2016. Girls enrolled in the remaining 59 schools are assigned to the Control group.

Data Collection

A baseline survey will be conducted for all 2,551 girls in the sample. The baseline survey will include a module on girls' time use and work history. In addition, the baseline survey will include a module that asks girls about their life skills using retrospective and prospective measures as well as a small number of objective measures designed to evaluate grit, perseverance, self-agency, and patience. A survey module conducted with the household head will collect supplementary information on household members' demographic characteristics, household income/expenditures, unexpected household shocks, etc. Finally, a survey module conducted with girls' caregivers will collect information on parental perceptions of girls' life skills and girls' expected educational achievement. The set of survey-based and task-based measures described above will be complemented by educational data collected from girls' schools that records student grades and attendance

history.[1] All of the measures described herein will be collected at baseline and again at the time of endline surveying. In addition, we intend to conduct a short cognitive test that will be used to objectively measure girls' academic performance; this will be collected at both endline, and at baseline subject to available funding.

Balance Check

Randomization balance will be established by comparing baseline data for girls in treatment and control schools.   To test for statistical balance, we will run a series of regressions of school and girl characteristics on an indicator variable characterizing the treatment assignment of the girl belonging to the relevant household and an indicator variable for randomization strata. Standard errors will be clustered at the school level and an F-test will be used to determine whether we can statistically reject the null hypothesis that the distribution of measures is the same for treatment and control students.

Estimating Treatment Effects

To identify the impacts of the Room to Read intervention, we will employ a benchmark specification that is an ordinary least squares (OLS) regression of each outcome of interest on an indicator variable for treatment assignment, an indicator variable for randomization strata, a vector of age dummies, indicators for 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 inclusion of the lagged control measure will serve to improve precision of estimated treatment effects as will the age and employment type controls. A small number of additional life skills questions may be ended at endline, at which point girls will be older and more able to answer questions (for example, related to future discounting and scenario-based behavior) that were deemed too challenging at baseline. In specifications examining these outcomes, we will control for lagged values of overall life skills indices since lagged outcome values will be unavailable. In all specifications, standard errors will be clustered at the school level.   Our sample will include 119 clusters.

We will examine the impact of the treatment on four primary sets of outcomes.

To deal with multiple inference concerns, we will employ the Benjamini-Hochberg correction method within each of these four categories to determine statistical significance of findings.

1) School progression and completion

_____

[1] Note: The inclusion of this complementary educational data is subject to governmental approval.

The key outcome measures here include school dropout, progression from one grade to the next, and school attendance. We hypothesize that the Room to Read intervention will reduce dropout, increase the probability of grade progression, and increase attendance rates.

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 eight based on response to household survey Q215. Variable will be coded as "0" for girls not enrolled in school or girls who have not advanced to class eight.

School attendance: This will be measured based on responses to child surveys Q205 and Q206. Two attendance measures will be constructed. The first will divide Q205 by Q206 responses, and the second will be an indicator for whether Q205 takes on a value greater than "0". Both variables will be coded as missing if Q206 is equal to "0".

In school progression, dropout and school attendance analysis, we propose to include additional variables that control for the following: age at enrollment (Q208 in household survey) and mother's educational attainment (Q213). We hypothesize that these control variables will serve as key predictors of the relevant outcomes.

We also propose to conduct heterogeneity analysis for the same set of outcomes based on child grade-for-age at baseline, mother's education, school quality, and baseline cognitive test performance. 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 child grade-for-age, the second specification will interact Treatment assignment with a discrete variable measuring mother's years of education, the third specification will interact Treatment assignment with a continuous variable measuring normalized school quality, and the fourth specification will interact Treatment assignment with a continuous variable measuring normalized cognitive test score. 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). In conducting this heterogeneity analysis, we will test for balance in the randomization within each subgroup and only report comparisons that are balanced. For the economic shock related outcomes, we will test for a correlation between treatment status and the prevalence of shocks in the last 12 months.

2) Life skills

Scaled scores for the three objective, task-based measures included in the survey will be used to characterize treatment effects. These three 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. Key survey-based measures used to evaluate program impacts will 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 Cantrill'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, and parental attitudes towards girl's marriage timing.

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 follow-up data when outcomes from the follow-up data are being analyzed).

We hypothesize that the Room to Read intervention will increase girl's patience based on measured future discounting, improve performance on mirror drawing and scavenger hunt tasks, and result in increases in the value of the indices described below (that are constructed based on both child and parent survey responses).

Future discounting: Responses to child survey Q510a-Q510f will be coded in order to generate a discrete variable that is set equal to the value corresponding to the first response in which the girl chooses present over future payout (there are six possible values).

Mirror drawing: Two variables will be constructed to measure task performance. The first 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 attempted mirror drawing tasks.

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).

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. Q507a-Q507d responses will take on one of two possible values, while Q508c-Q508g will take on one of three possible values. "Don't Know" responses to Q507a-Q507d will be coded as missing.

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.

Cantrill'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.

We also propose to conduct heterogeneity analysis for the same set of outcomes discussed above based on child grade-for-age at baseline, mother's education, school quality, baseline cognitive test performance, and indicators for whether household has experienced various types of household shocks (as identified in Q111 and Q113).

In conducting this heterogeneity analysis, we will test for balance in the randomization within each subgroup and only report comparisons that are

balanced. For the economic shock related outcomes, we will test for a correlation between treatment status and the prevalence of shocks in the last 12 months.

    3) Child labor

Outcomes of interest include participation in hazardous child labor, economic activity both inside and outside the home, and detailed information on time allocation. Data will be collected to serve as proxies for bonded labor and human trafficking, but we do not expect prevalence rates that would permit further analysis on bonded labor or human trafficking.

We hypothesize that the Room to Read intervention will reduce time spent working and incidence of forced labor.

Working Child
- Children who have worked in the 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 either 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.

Hours 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.

We also propose to conduct heterogeneity analysis for the same set of outcomes based on the child's age at baseline, based on the type of employment most important to the child's household at baseline (question 104 of the household survey), based on the child's land ownership at baseline (question 106 of the household survey), based on whether the household has affected by any general problems in the last 12 months (question 111 of the household survey), based on whether the household has experienced any serious illnesses in the last 12 months (question 113, parts D-F), and based on whether the household has experienced any household problems that resulted in the loss of household income or inheritances (questions 113 and 114 of the household survey). Each source of heterogeneity will be considered separately. In addition to including an indicator for Treatment assignment and the type of heterogeneity being examined, heterogeneity will be examined by interacting indicator(s) for the type of heterogeneity described above and the treatment assignment. In conducting this heterogeneity analysis, we will test for balance in the randomization within each subgroup and only report comparisons that are balanced. For the economic shock related outcomes, we will test for a correlation between treatment status and the prevalence of shocks in the last 12 months.

All specifications will control for the child's age at baseline with indicators as well as indicators for the most important type of employment in the household at baseline.

4) Cognitive skills and academic achievement

The key outcomes of interest here will include a continuous measure of student grade point average collected from administrative data, time spent studying, and performance on a cognitive test administered by the research team at both baseline and endline (subject to available funding).

We hypothesize that the Room to Read intervention will increase time spent studying.

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.

We also propose to conduct heterogeneity analysis for the same set of outcomes discussed above based on child grade-for-age at baseline, mother's education, school quality, and indicators for whether household has experienced various types of household shocks (as identified in Q111 and Q113.

In conducting this heterogeneity analysis, we will test for balance in the randomization within each subgroup and only report comparisons that are balanced. For the economic shock related outcomes, we will test for a correlation between treatment status and the prevalence of shocks in the last 12 months.

Survey Attrition

By collecting contact information for each household at baseline and by engaging in multiple follow-up visits in cases in which respondents are not initially reached by enumerators, we hope to minimize survey attrition. Nonetheless, we will test for differential attrition by treatment assignment at the time of endline analysis and will present estimates using an appropriate bounding procedure if differential attrition is detected.

Questions with Limited Variation

In order to limit noise caused by variables with minimal variation, questions for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis and will not be included in any indicators or hypothesis tests.

Missing data from item non-response

Unless explicitly stated above, there will be no imputation for missing data due to item non-response at endline. Missing data on baseline variables will be dummied out of the relevant specifications.

Power Calculations

Given the available sample, power calculations suggest that we will be able to detect increases of 4 percentage points in child enrollment and decreases of 3 and 8 percentage points for child marriage and wage work, respectively. These minimum detectable effects sizes are calculated based on 0.90 power, a one-sided hypothesis test conducted with a p-value of .05, and assumed intraclass correlations of 0.06, 0.00, and 0.10 for child enrollment, child marriage and wage work, respectively. We note that intraclass correlation values are calculated based on existing pilot data to the extent possible.

Appendix One: Banned Occupations for Children Below the Age of 18, 1986 Child Labor Act including all subsequent amendments

(1) Transport of passengers, goods or mails by railway;

(2) Cinder picking, clearing of an ash pit or building operation in the railway premises;

(3) Work in a catering establishment at a railway station, involving the movement of a vendor or any other employee of the establishment from one platform to another or into or out of a moving train;

(4) Work relating to the construction of a railway station or with any other work where such work is done in close proximity to or between the railway lines;

(5) A port authority within the limits of any port;

(6) Work relating to selling of crackers and fireworks in shops with temporary licences;

(7) Abattoirs/slaughter Houses;

(8) Automobile workshop and garages;

(9) Foundries;

(10) Handling of toxic or inflammable substances or explosives;

(11) Handloom and powerloom industry;

(12) Mines (underground and underwater) and collieries;

(13) Plastic units and fibreglass workshops;

(14) Employment of children as domestic workers or servants;

(15) Employment of children in dhabas (road side eateries), restaurants, hotels, motels, tea shops, resorts, spas or other recreational centres;

(16) Diving.

Appendix Two: Banned Processes for Children Below the Age of 18, 1986 Child Labor Act including all subsequent amendments

(1) Bidi-making.

(2) Carpet-weaving including preparatory and incidental process thereof.

(3) Cement manufacture, including bagging of cement.

(4) Cloth printing, dyeing and weaving including processes, preparatory and incidental thereto.

(5) Manufacture of matches, explosives and fire-works.

(6) Mica-cutting and splitting.

(7) Shellac manufacture.

(8) Soap manufacture.

(9) Tanning.

(10) Wool-cleaning.

(11) Building and construction industry including processing and polishing of granite stones.

(12) Manufacture of slate pencils (including packing).

(13) Manufacture of products from agate.

(14) Manufacturing processes using toxic metals and substances such as lead, mercury, manganese, chromium, cadmium, benzene, pesticides and asbestos.

(15) 'Hazardous process' as defined in section 2 (cb) and 'dangerous operation' as notified in rules under section 87 of the Factories Act, 1948 (63 of 1948).

1. Ferrous Metallurgical Industries

- Integrated Iron and Steel
- Ferrow-alloys
- Special Steels

2. Non-ferrous metallurgical Industries

   - Primary Metallurgical Industries, namely, zinc, lead, copper, manganese and aluminium

3. Foundries (ferrous and non-ferrous)

   - Castings and forging including cleaning or smootherning/roughening by sand and shot blasting

4. Coal (including coke) industries

   - Coal , Lignite, Coke, etc.
   - Fuel Gases (including Coal Gas, Producer Gas, Water Gas)

5. Power Generating Industries

6. Pulp and paper (including paper products) industries

7. Fertiliser Industries

   - Nitrogenous
   - Phosphatic
   - Mixed

8. Cement Industries

   - Portland Cement (including slag cement, puzzolona cement and their products)

9. Petroleum Industries

   - Oil Refining
   - Lubricating Oils and Greases

10. Petro-chemical Industries

11. Drugs and Pharmaceutical Industries

    - Narcotics, Drugs and Pharmaceuticals

12. Fermentation Industries (Distilleries and Breweries)

13. Rubber (Synthetic) Industries

14. Paints and Pigment Industries

15. Leather Tanning Industries

16. Electro-plating Industries

17. Chemical Industries

- Coke Oven by-products and Coaltar Distillation products
- Industrial Gases (nitrogen, oxygen, acetylene, argon, carbon dioxide, hydrogen, sulphur dioxide, nitrous oxide, halogenated hydrocarbon, ozone, etc.)
- Industrial Carbon
- Alkalies and Acids
- Chromates and dichromates
- Leads and its compounds
- Electrochemicals (metallic sodium, potassium and magnesium, chlorates, perchlorates and peroxides)
- Electrothermal produces (artificial abrasive, calcium carbide)
- Nitrogenous compounds (cyanides, cyanamides and other nitrogenous compounds)
- Phosphorous and its compounds
- Halogens and Halogenated compounds (Chlorine, Flourine, Bromine and Iodine
- Explosives (including industrial explosives and detonators and fuses)

18. Insecticides, Fungicides, Herbicides and other Pesticides Industries

19. Synthetic Resin and plastics

20. Man-made Fibre (Cellulosic and non-cellulosic) Industry

21. Manufacture and repair of electrical accumulators

22. Glass and Ceramics

23. Grinding or glaxing of metals

24. Manufacture, handling and processing of asbestos and its products

25. Extraction of oils and fats from vegetable and animal sources

26. Manufacture, handling and use of benzene and substances containing benzene

27. Manufacturing processes and operations involving carbon disulphide

28. Dyes and Dyestuff including their intermediates

29. Highly flammable liquids and gases

(16) Printing as defined in section 2(k) (iv) of the Factories Act, 1948 (63 of 1948).

(17) Cashew and cashewnut descaling and processing.

(18) Soldering processes in electronics industries.

(19) 'Aggarbatti' manufacturing.

(20) Automobile repairs and maintenance including processes incidental thereto, namely, welding, lathe work, dent beating and painting.

(21) Brick kilns and roof titles units.

(22) Cotton ginning and processing and production of hosiery goods.

(23) Detergent manufacturing.

(24) Fabrication workshops (ferrous and non-ferrous).

(25) Gem cutting and polishing.

(26) Handling of chromite and managanese ores.

(27) Jute textile manufacture and coir making.

(28) Lime kilns and manufacture of lime.

(29) Lock making.

(30) Manufacturing processes having exposure to lead such as primary and secondary smelting, welding and cutting of lead-painted metal constructions, welding of galvanized or zinc silicate, polyvinyl chloride, mixing (by hand) of crystal glass mass, sanding or scrapping of lead paint, burning of lead in enamelling workshops, lead mining, plumbing, cable making, wire patenting, lead casting, type founding in printing shops. Store type setting, assembling of cars, shot making and lead glass blowing.

(31) Manufacture of cement pipes, cement products and other related work.

(32) Manufacturing of glass, glassware including bangles, flourescent tubes, bulbs and other similar glass products.

(33) Manufacture of dyes and dye stuff.

(34) Manufacturing or handling of pesticides and insecticides.

(35) Manufacturing or processing and handling of corrosive and toxic substances, metal cleaning and photo engraving and soldering processes in electronic industry.

(36) Manufacturing of burning coal and coal briquettes.

(37) Manufacturing of sports goods involving exposure to synthetic materials, chemicals and leather.

(38) Moulding and processing of fibreglass and plastic.

(39) Oil expelling and refinery.

(40) Paper making.

(41) Potteries and ceramic industry.

(42) Polishing, moulding, cutting, welding and manufacture of brass goods in all forms.

(43) Process in agriculture where tractors, threshing and harvesting machines are used and chaff cutting.

(44) Saw mill—all processes.

(45) Sericulture processing.

(46) Skinning, dyeing and processes for manufacturing of leather and leather products.

(47) Stone breaking and stone crushing.

(48) Tobacco processing including manufacturing of tobacco, tobacco paste and handling of tobacco in any form.

(49) Tyre making, repairing, re-treading and graphite benefication.

(50) Utensils making, polishing and metal buffing.

(51) 'Zari' making (all processes).

(52) Electroplating.

(53) Graphite powdering and incidental processing.

(54) Grinding of glazing of metals.

(55) Diamond cutting and polishing.

(56) Extraction of slate from mines.

(57) Rag picking and scavenging.

(58) Processes involving exposure to excessive heat (e.g. working near furnace) and cold.

(59) Mechanised fishing.

(60) Food processing.

(61) Beverage industry.

(62) Timber handling and loading.

(63) Mechanical Lumbering.

(64) Warehousing.

(65) Processes involving exposure to free silica such as slate, pencil industry, stone grinding, slate stone mining, stone quarries, agate industry.

Section B: Addendum to Pre-Analysis Plan (August 2018)

After some field experience, we modified three types of data collection and amended the AEA registration entry in August 2018 while the endline survey was fielding. The first change was the addition of three non-cognitive indices to the endline survey: the Rotter locus of control module, a module focusing on perceived stress, and the Rosenberg self-esteem module. The second change was the specification of the cognitive test to be employed at endline; the ASER test was employed, following extensive exploration of appropriate instruments as well as piloting. The third change related to the measurement of future discounting. Following 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. In addition, we clarified in the addendum how regression models would be specified for study participants with missing baseline outcome measures. The addendum to the pre-analysis plan is included below in full.

Additional life skills modules

Three additional modules will be added to the life skills section of the endline survey: 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. We will follow scoring instructions provided for each module. Then, we will standardize the final score relative to the control population as described for the other life skills measures in the original analysis plan. These additions will generate three new non-cognitive indices to be analyzed in conjunction with the non-cognitive indices postulated at baseline: an index of locus of control, an index of perceived stress, and an index of self-esteem.
The analysis will note that these outcome variables were added prior to endline, rather than at baseline. The prospect of adding additional life skills modules was noted in our analysis plan. Noted in that original analysis plan is that we will include baseline values of our other life skills indexes as controls in looking at the impact of treatment on these new modules.

Cognitive skills and academic achievement – more detailed measurement

Our original analysis plan noted that the fourth outcome of interest was cognitive skills and academic achievement. The key outcomes of interest here will include continuous measure of student grade point average, and performance on the cognitive test administered by the research team in conjunction with local schools. We can now add further details about the cognitive test that will be administered at endline. We will be employing the ASER tests in mathematics, Hindi and English

utilized by Pratham, a major educational non-governmental organization in India. The ASER tests are regularly administered across India in both rural and urban settings for students of the same target age group employed in this evaluation. The ASER instrument, including instructions for how it is scored, are attached as an appendix to this document. The research team will be utilizing the same protocols developed by Pratham in order to score each child at a given level of proficiency in each subject. For mathematics, the four levels of proficiency are recognition of single digit numbers; recognition of double digit numbers; subtraction; and division. For reading in both Hindi and English respectively, the four levels are reading letters, reading words, reading paragraphs, and reading stories.

Our original analysis plan did not stipulate how we would analyze the ASER scores. We will mimic our approach used for other outcomes. However, in the ASER score case, we lack baseline values. We propose to 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 school-level administrative data. (The school-level administrative data was collected from school records by the research team.)

<u>Sample girls with missing baseline values</u>

As noted in our previous documentation, we were able to complete only part of the baseline survey for a small number of sampled girls. 16 girls had baseline child surveys completed, and no baseline household survey. 44 girls had baseline household surveys completed, and no baseline child survey. These patterns reflect either the absence of the household member required to complete the household survey, or the household head's unwillingness to provide consent for a survey of the child.

In addition, there are 16 sample girls corresponding to a single primary school (Arain) who were not surveyed at baseline, as another school was erroneously identified as the treatment school. The girls from the correct treatment school are represented in follow-up data collection but not in the baseline survey.

In analysis, sample girls with missing baseline data will be included in all regressions, and the specification will add dummy variables equal to one for any observations missing baseline data.