

Primary Outcome

The primary outcome of interest is an index of indicators for *having experienced various forms of sexual violence since baseline*. Using the answers to whether the individual has experienced each of the following traumatic events since baseline (i.e. during the past 24 months), we construct an index of sexual violence using Inverse Covariance Weighting (ICW): non-consensual touching, pressure to have sex, attempted rape, and rape. [Q1-Q4 in the participant survey]

To construct this index and other indices throughout the study using ICW, we followed Casey et al. (2012), which consists of the following steps: (i) standardizing each of index components using the control group's mean and standard deviation; (ii) computing the variance-covariance matrix of the standardized variables; (iii) generating a weighted average of the standardized variables where the weights are proportional to the sums of the rows of the inverted variance-covariance matrix; and (iv) standardizing this weighted average using the control group's mean and standard deviation.

In addition to the primary outcome defined as described above, we also report whether the respondent experienced physical violence.

Secondary Outcomes

We identified six secondary outcomes: schooling, mental health, gender norms, life skills, sexual and reproductive history, and protective factors. Questions in each of these domains are aggregated into indices, which were constructed using Inverse Covariance Weighting (ICW) as described above.

In the construction of all seven indices (one primary plus six secondary outcome domains), observations with refusals (meaning an entire sub-component or scale is missing) or item non-response (meaning that some but not all questions within a sub-component or scale are missing) are dropped from the analysis. In Appendix A, we check the robustness of our impact findings to how we treat item non-response during index creation.

Below, we describe the components of each of the six indices in detail.

Schooling

The schooling index is composed of two variables:

- *Highest grade completed*, a discrete variable that captures the highest grade the respondent has completed [B2 in the participant survey], and
- *School enrollment* in academic year 2016-17, an indicator variable equal to 1 if the individual was enrolled in school during the reference period. [B6 in the participant survey]

Psychological wellbeing

The psychological wellbeing index has three components:

- Rosenberg self-esteem scale,
- Short Mood and Feelings Questionnaire (SMFQ), and
- Children's Revised Impact of Event Scale (CRIES8).

All sub-components of the index, i.e. the Rosenberg scale, the SMFQ and the CRIES8, have been constructed such that higher values imply better outcomes.

The *Rosenberg self-esteem scale* is composed of 10 questions, each scored on a scale of 0-3, for a total of 30 possible points. Scores above 15 may indicate low self-esteem¹. [F1-F10 in the participant survey]

The *SMFQ* is composed of 12 questions, each scored 0-2, for a total of 24 points. This instrument is used as a screening tool for situations where depression is suspected.² [H1-H6 and H8-H13 in the participant survey]³

The *CRIES8* measure is composed of eight questions, each scored on a scale of 0-5, for a total of 40 possible points. This measure is a screening tool used to measuring children at risk for post-traumatic stress symptoms⁴. In the endline survey, the CRIES8 questions were only administered to individuals who reported experiencing physical violence (P1 in the participant survey) or sexual violence (Q1-Q4 in the participant survey) at either baseline or endline. Those who did not report ever experiencing these traumatic events were assigned the highest possible score of 40, with higher scores indicating better outcomes. [R1-R8 in the participant survey]

Gender norms

The gender norms index has two components: the gender equity score and an index of attitudes towards intimate partner violence (IPV). All components of the gender equity and the violence-unjustified indices have been constructed such that higher values imply better outcomes (more gender equitable and less tolerant towards IPV, respectively). Scores are summed over all binary items within each index and observations with missing values are handled as described earlier (see the sub-section on **Psychological Wellbeing** above).

The *gender equity* index is made up of five statements, with each of which the respondent must agree or disagree, scored 0 or 1. The score, which is equal to the sum of five items, is standardized, using the mean and standard deviation in control group. [J2-J6 in the participant survey]

The *violence unjustified* index (i.e. *attitudes towards IPV*) is comprised of six statements asking the respondent whether the husband is justified in beating his wife in different hypothetical scenarios, scored 0 or 1. The score is standardized, which is equal to the sum of six items, using the mean and standard deviation in control group. [J7a-J7f in the participant survey]

¹http://fetzer.org/sites/default/files/images/stories/pdf/selfmeasures/Self_Measures_for_Self-Esteem_ROSENBERG_SELF-ESTEEM.pdf

² <http://sspeditricassociates.com/Forms-and-Policies/Forms/Behavioral,-Mental-Health-Assessment-Forms/Short-Mood-Form-Child-and-Parent.aspx>

³ Note the SMFQ is normally comprised of 13 items. However, one item (“in the past two weeks, I found it hard to think properly or concentrate”) was omitted from the endline survey by a programmer mistake.

⁴ <http://www.childrenandwar.org/measures/children's-revised-impact-of-event-scale-8---cries-8/>

Life skills

The *life skills* index has five components: HIV knowledge, health knowledge, financial literacy, knowledge of condom effectiveness, and healthy intimate (heterosexual) relationships. All components of each score have been constructed such that higher values imply better outcomes. Scores are summed over all items within each index and observations with missing values are handled as described earlier (see the sub-section on **Psychological Wellbeing** above). All index components reported are standardized.

The *HIV knowledge* index is made up of seven “true or false” questions, scored 1 or 0 for correct or incorrect answers, respectively. This group of questions was only administered to those who answered in the affirmative to question (M1): “Have you heard of HIV or AIDS?” In case the respondent had not heard of HIV, these questions would not be asked. The index is constructed on the extensive margin, i.e. for everyone, by replacing the missing answers from respondents who have never heard of HIV with the wrong answer for each of the seven statements. The score, which is equal to the sum of seven items, is standardized, using the mean and standard deviation in control group. [M2-M8 in the participant survey]

The *health knowledge* index is composed of four questions, scored for a maximum of 19 points. K1 is a “true or false” statement, asking whether the monthly menstrual period is something to be ashamed of, which is scored 0 if true and 1 if false. The other questions in this index allow multiple correct answers. We have constructed the score by counting how many of the correct options the respondent had selected without prompting. For example, question K2 asks: “As girls grow into women, what changes happen in their bodies? Can you name a few?” The question has six options, and the respondent’s score is the total number of options mentioned. The score, which is equal to the sum of scores from these four items, is standardized, using the mean and standard deviation in control group. [K1, K2, K8, and L1a in the participant survey]

The *financial literacy* index is comprised of eight questions, scored for a maximum of 20 points. The questions collect information on the individual’s savings, expenditures, and financial knowledge. The items are a combination of binary indicators (yes/no) and multiple correct answers. For the latter, the number of right options mentioned by the respondent is counted. The score, which is equal to the sum of scores from these eight items, is standardized, using the mean and standard deviation in control group. [I1-I8 in the participant survey]

The *knowledge of condom effectiveness* score is composed of five true or false statements, with correct/incorrect answers assigned a 1/0, for a total of five points. The score is standardized, using the mean and standard deviation in control group. [N2-N6 in the participant survey]

The last component of the life skills index is the *healthy intimate (heterosexual) relationships* index. This is comprised of one “select multiple answers” question: “What describes a good, healthy intimate male-female relationship, like between husband and wife, or between girlfriend and boyfriend?” To construct the score, we selected 9 of the 10 options, each of which represent (in our opinion) a healthy aspect of an intimate male-female relationship and counted how many of these the respondent mentioned without being prompted. We did not consider financial support to be what describes a healthy intimate male-female relationship and, as such, did not include it in the score if the respondent chose that option. The score is standardized, using the mean and standard deviation in control group. [J1a in the participant survey]

Sexual history and reproductive health

This index has five components: three indicator variables for: never married, never had sex and never pregnant, a discrete variable for the number of partners in the past 12 months, and a safe sex index regarding condom use [*O1, O10, O23, O14, and O15, respectively, in the participant survey*]. The safe sex index is standardized.

All components of the index have been constructed such that higher scores lower levels of sexual activity, marriage, and pregnancy (as the subjects are 15-16 years old at endline), as well as safer sex. Observations with missing values for any of these five components are dropped from the analysis.

Never married, never had sex, and never pregnant are indicator variables. *Never had sex* is the gateway question for the following questions: ever pregnant, number of partners, and condom use, which were only asked to those respondents who reported having ever had sex. *Never pregnant, number of partners, and the safe sex index* have been constructed on the extensive margin – with respondents who reported having never had sex assumed to be never pregnant, having no sexual partners, and assigned the highest score in the safe sex index.

The *safe sex index* is composed of two questions: “In the past 12 months, how often did you or your partners use a condom (male or female) during man and woman business?” and “The last time you did man and woman business, did you use a condom?” The former question was transformed into an indicator variable equal to 1 if the respondent said “always.” The score, which is equal to the sum of scores from these two items, is standardized, using the mean and standard deviation in control group.

Protective factors

The *protective factors* index includes questions both from the individual primary respondent survey and from the caregiver one. It has three components: a social capital score, a caregiver gender norms score and a caregiver child rearing score. All components of the index have been constructed such that higher scores imply more desirable outcomes. Scores are summed over all items within each index and observations with missing values are handled as described earlier (see the sub-section on **Psychological Wellbeing** above).

The social capital index is composed of nine questions for a maximum score of 14. It consists of questions that inquire about whether the respondent has a big person in their life, whether they have a safe place to go and whether they have friends their age, etc. Questions D1-D3 are scored on a scale of 0-2, while D4, E1, E5, E6, and Q10 are indicator variables. Question E7, “How many girlfriends of your age do you have outside the household?” has been categorized to be equal to 0 if none, 1 if 1-2, 2 if 3-4, and 3 if the respondent has more than 4 friends. The score, which is equal to the sum of scores from these nine items, is standardized, using the mean and standard deviation in control group. [*D1-D4, E1, E5-E7 in the participant survey*]

The gender norms score is composed of 11 true or false statements asked to the primary caregiver of the young female study participant. These questions aim at capturing the attitude of caregivers towards their daughters by asking the respondent whether they agree with statements such as the following: “It is important that sons have more education than

daughters,” or “Daughters should be sent to school only if they are not needed to help at home.” The score, which is equal to the sum of scores from these 11 items, is standardized, using the mean and standard deviation in control group. [C1-C11 in the primary caregiver survey]

Finally, the *child rearing score* is comprised of seven questions asked to the primary caregiver. These questions ask caregivers about their hopes for their daughters in terms of their education, marriage, fertility, and labor market participation. D1, D2 and D7 are indicator variables; D3 is transformed into an indicator variable that equals 1 if the caregiver hoped the girl would complete high school; D4 is transformed into an indicator variable that equals 1 if the caregiver hoped the girl would stay in school at least until the age of 18; and another indicator variable takes the value of 1 if the caregiver wished for the girl to stay unmarried (D5) and not pregnant (D6) until at least the age of 18 and 0 otherwise. The score, which is equal to the sum of scores from these six items, is standardized, using the mean and standard deviation in control group. [D1-D7 in the primary caregiver survey]

Analysis

The analysis follows standard guidelines and operating procedures for the analysis of RCTs as described in Bruhn and McKenzie (2009) and Lin, Green, and Coppock (2015). We envision 11 main tables: baseline balance, attrition, effect of cash transfers on program take-up (GE vs. GE+), programs impact on each of the seven primary and secondary outcomes (and their subcomponents), and a summary table that includes only the seven primary and secondary outcomes along with p-values and false discovery rate-adjusted q-values for treatment effects in GE and GE+.

Table 1: Baseline Balance

We include eight baseline covariates in this table: age and the seven primary and secondary outcome indices. For each variable, we report the mean in the control group, the differences in GE and GE+ with the control group, the p-value of the difference, and the number of observations, using a standard linear regression model with clustered standard errors. We also report a “Chi-Squared Test for the Joint Orthogonality of all eight baseline covariates,” using a multinomial logit regression with treatment status as the dependent variable and the baseline covariates as the independent variables. Each of the seven indices presented in this table has been constructed as the average of non-missing index sub-components, which maximizes sample size, then standardized using the control group mean and standard deviation. An additional row of joint orthogonality tests is presented for an alternative method of constructing the seven indices, where observations with item non-response for a sub-component are assigned a missing value for the entire index.

Balancing of Variables across Groups - Baseline					
Variable	Mean (standard deviation) for control group	Difference in Means (with the control group)		Test for Equality of Parameters (p-values)	Number of Observations
		GE	GE+	GE=GE+	
	(1)	(2)	(3)	(4)	(5)
Age					
Schooling					
Psychosocial					
Gender equality					
Life skills					
Sexual history					
Sexual violence					
Social capital					
Chi-Squared Test for Joint Orthogonality of All Variables (p-value)					

Notes: For 'Difference in Means' columns, stars are used to represent statistical significance according to: .01 - ***, .05 - **, .1 - *. Numbers in parentheses are standard errors, except for 'Mean for control group' column, where they are standard deviations.

Table 2: Attrition

Attrition analysis is also standard. A binary indicator for whether an individual was lost to follow-up will be regressed on the same eight baseline covariates reported in the baseline balance table, with each of these variables fully interacted with indicators for GE and GE+ as follows:

$$Y_{ij} = \alpha + \beta X_{ij} + \gamma^2 T_j^2 + \gamma^3 T_j^3 + \delta^2 T_j^2 X_{ij} + \delta^3 T_j^3 X_{ij} + Z_j + \varepsilon_{ij},$$

where Y_{ij} is an indicator for being lost to follow-up for individual i in cluster j , T_j^2 and T_j^3 are binary indicators for GE and GE+, X_{ij} is the vector of eight standardized baseline covariates reported in the baseline balance table, and Z_j are block fixed effects. Missing baseline covariates are replaced with the sample mean in this analysis, following standard operating procedures suggested by Lin, Green, and Coppock (2015).

Mean attrition rate in the control group at the 24-month follow-up will be reported, along with joint F-tests for baseline covariates and interactions for each treatment arm.

In case of significant differences in attrition between study arms in either levels or in baseline characteristics (as indicated by the joint F-test of interactions for δ^2 and δ^3), we will report the robustness of the findings in the main summary table using inverse probability weighting, Lee bounds, and Manski bounds.

TABLE 2: Attrition Individual Level

		Dependent Variable: Binary Indicator for Girl Lost to Follow-Up	
		24-Month Follow-Up	
Variable		(1)	(2)
GE			
GE+			
Girl variables	Age		
	Schooling Index		
	Psychological Wellbeing Index		
	Gender Equality Index		
	Life Skills Index		
	Sexual and Reproductive History Index		
	Sexual Violence Index		
	Social Capital Index		
Interactions: GE Dummy x Girl variables	GE x Age		
	GE x Schooling		
	GE x Psychosocial		
	GE x Gender equality		
	GE x Life skills		
	GE x Sexual history		
	GE x Sexual Violence		
	GE x Social capital		
Interactions: GE+ Dummy x Girl variables	GE+ x Age		
	GE+ x Schooling		
	GE+ x Psychosocial		
	GE+ x Gender equality		
	GE+ x Life skills		
	GE+ x Sexual history		
	GE+ x Sexual Violence		
	GE+ x Social capital		
Mean (standard deviation) of dependent variable for the control group			
Joint F-test of Baseline Controls (minus interactions) - p-value			
Joint F-test of Interactions - p-value		with GE:	
		with GE+:	
Number of observations			

Notes: .01 - ***; .05 - **, .1 - *; Regressions are at the individual level, standard errors are clustered by village. All test scores are standardized by using means and standard deviations from the control group at baseline. Standard errors in parentheses. We replaced the missing values of index variables with their averages at baseline for the overall sample. Please see appendix for index components.

Table 3: Impact of conditional cash transfers on program take-up

To estimate intention-to-treat (ITT) effect of providing small cash transfers conditional on attending each GE session (GE + cash), we employ a regression model of the following form at endline only among those in GE and GE+, i.e. excluding the control group:

$$Y_{ij} = \alpha + \gamma^2 T_j^3 + \varepsilon_{ij},$$

where Y_{ij} is the number of program, sessions attended by individual i (adolescent female respondent or her primary caregiver) in cluster j ; T_j^3 is a binary indicator for the cluster-level intervention GE+. The regressions also absorb the blocks used for random assignment. The standard errors ε_{ij} , clustered at the town/village level, account for both the design effect of the cluster-level treatment and heteroskedasticity inherent in the regression model.

Tables 4-10: Impacts on each of the seven primary and secondary outcomes

To estimate intention-to-treat (ITT) effects of each intervention on child outcomes, we employ a regression model of the following form at the 24-month follow-up data collection:

$$Y_{ij} = \alpha + \gamma^2 T_j^2 + \gamma^3 T_j^3 + \beta X_{ij} + \varepsilon_{ij},$$

where Y_{ij} is an outcome variable for individual i in cluster j ; T_j^2 and T_j^3 are binary indicators for cluster-level interventions GE and GE+, respectively; and X_{ij} is a vector of baseline covariates consisting of the lagged dependent variable and age of the individual in years. The regressions also absorb the blocks used for random assignment. The standard errors ε_{ij} , clustered at the town/village level, account for both the design effect of the cluster-level treatment and heteroskedasticity inherent in the regression model.

To show the reader if there are any effects in the combined treatment group, which will have more statistical power, we will also include a row in each table for “Any Treatment” in addition to GE and GE+.

TABLE 4: Impacts on Adolescent Girls Outcomes - 24-Month Follow-Up

	Dependent Variable:					
	Domain					
	Main Index		Components		Component	
	(1)	(2)	(3)	(4)	(5)	(6)
GE						
GE+ (GE + cash)						
Any Treatment (GE or GE+)						
Mean and Standard Deviation of un standardized dependent variable in the control group						
F-test for Equality of Parameters (p-value)	GE=Geplus					
Block Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Dependent Variable/ Age	No	Yes	No	Yes	No	Yes
Number of observations						

Table 11: False Discovery Rate-adjusted Q-values the seven primary and secondary outcomes

False Discovery Rate-adjusted Q-values for each pairwise comparison, which is the smallest level of statistical significance at which the null hypothesis is rejected, will be calculated using the Benjamini and Hochberg (1995) procedure as described in Anderson (2008) and presented in this table along with the original effect sizes and standard errors from the original impact table.

Panel A will present corrections for 21 hypothesis tests (7 outcomes by 3 pairwise comparisons of GE and GE+ with the control group), while Panel B will present FDR-adjusted Q-values for 7 hypothesis tests (7 outcomes with only one pairwise comparison of “Any Treatment” to the Control Group).

In addition to these tables, there may be a table for robustness checks (such as robustness to different assumptions regarding observations lost to follow-up) and a table discussing the heterogeneity of impacts. We will examine the heterogeneity of effects at the 24-month follow-up, using the following regression model:

$$Y_{ij} = \alpha + \beta X_{ij} + \gamma^2 T_j^2 + \gamma^3 T_j^3 + \delta^2 T_j^2 X_{ij} + \delta^3 T_j^3 X_{ij} + \varepsilon_{ij},$$

where Y_{ij} is an outcome variable for individual i in cluster j , T_j^2 and T_j^3 are binary indicators for cluster-level interventions GE and GE+, and X_{ij} is a vector of centered baseline covariates. Coefficients γ^2 and γ^3 are estimates of ITT effects for GE and GE+, while δ^2 and δ^3 are estimates of heterogeneity of treatment effects. The regression will again absorb the block fixed effects.

In matters related to the handling of data, such as how to deal with missing covariates, etc., we will follow the standard operating procedures for the analysis of RCTs in Lin, Green, and Coppock (2015). The paper will also closely follow the style of analysis in Özler et al. (2017).

References

Bruhn, M. and McKenzie, D. (2009). “In pursuit of balance: Randomization in practice in development field experiments,” *American Economic Journal: Applied Economics*, 1(4): 200-232.

Lin, W., Green, D. P., and Coppock, A. (2016). “Standard Operating Procedures for Don Green’s lab at Columbia,” latest version available at: http://alexandercoppock.com/Green-Lab-SOP/Green_Lab_SOP.pdf

Özler, B., Fernald, L. C. H., Kariger, P., McConnell, C., Neuman, M., and Fraga, E. (2017). “Combining Preschool Teacher Training with Parenting Education: A Cluster-Randomized Controlled Trial,” latest version available at: <https://sites.google.com/site/decrgrberkozler/papers-by-date>