Girls' Economic Empowerment – A Randomized Experiment in Tanzanian Schools

Pre-Analysis Plan

Outline:

The timeline of the present project is as follows:

- Baseline study: March-April, 2013 (questionnaire attached)
- Intervention: June-September, 2013 (detailed information attached)
- Short term follow-up study: September-October, 2013 (questionnaire attached)
- Long term follow-up study: September-October, 2014 (planned)

This pre-analysis plan has been written prior to data analysis of the follow-up data and pre-commits the authors to defined specifications for estimating impacts.[1][2] It was written by Kjetil Bjorvatn, Lars Ivar Oppedal Berge, Vincent Somville and Bertil Tungodden.

1. Abstract

Teenage pregnancies are common in many low-income countries, but the reasons for why teenage girls become pregnant are not well understood. This is particularly unfortunate in a low-income context where pregnancy among young girls can result in illegal and often dangerous abortions, the spread of HIV/AIDS and, more generally, poor health, educational, and economic outcomes.

The present study takes place in Tanzania and investigates the fertility and economic decisions of girls when they are on the verge of making two of the most important decisions in their lives: what to do when leaving school and whether to start childbearing.

The present project investigates whether early childbearing reflects a lack of empowerment among young girls in Tanzania, with a focus on two different empowerment strategies. First, we offer an *information* treatment where young girls are targeted with detailed and extensive information about reproductive health, gender equality, and rights. Second, we offer an *opportunity* treatment where young girls are targeted with entrepreneursHIP training to improve their skills on how to run a business. Both treatments aim at empowering girls, but through different channels. The information treatment represents in many ways the classical approach in the field, where the underlying idea is that teenage pregnancies reflect lack of relevant information and personal control. The opportunity treatment, on the other hand, investigates whether teenage pregnancies to a large extent reflect a lack of economic opportunities. By comparing these two treatments, the study will provide novel insights on the relative importance of providing information and opportunities to adolescent

[1] We however acknowledge that, due to an administrative mistake, we had access to the short-term follow-up data on 4 variables in 3 schools before the registration of this plan was completed. This did not influence the writing of this plan in any manner. Note also that two other researchers involved in this project, Tausi Kida and Linda Helgesson Sekei, were running a qualitative component of the project between the interventions and the writing of this plan and were therefore excluded from the set-up of this plan.

[2] In the event that the structure of the data or other unforeseen factors necessitate adjustments in the methodology and specifications to be employed in the analysis, such adjustments will be documented with reference to the original specifications in this pre-analysis plan and accompanied by a justification of why such adjustments were necessary.

girls, which is important both from a theoretical and a policy perspective. In addition, since there may be important complementarities in how the two treatments work, we also offer a cross-treatment to a subsample of the girls.

An ultimate goal of this research project is to inform the design of sound policies, and we are thus particularly careful to *develop a cost-effective intervention with scaling-up potential:* all treatments are evaluated and compared not only in terms of impact but also on their relative economic costs. The interventions are based on current practices of our partners, which makes it easier to scale up the successful parts of the project and, more generally, to use the knowledge from the project to improve existing policies.

2. Interventions

Together with our partner Femina HIP, a leading NGO on reproductive health in Tanzania, we have decided to focus on girls who are in their last year of secondary school (Form IV in Tanzania). They are in the age interval where we observe a sharp increase in fertility. Most of them will not have access to further schooling, and will thus have to consider other opportunities, including opening a small scale business, which is a very common activity in the areas of this study.

The interventions consist in training sessions offered in a classroom setting on a weekly basis for eight weeks:

1) Reproductive health information treatment: An information course on reproductive health, gender equality, and rights.

Femina HIP, together with the research team, designed a tailored information program for the targeted girls, which provided both practical and objective information about reproductive health, such as information about contraception and the consequences of risky sexual behavior. In addition, the course focused on gender equality, rights, and women's empowerment.

2) Economic opportunity treatment: EntrepreneursHIP training to improve economic opportunities.

Femina HIP, together with the research team, designed a tailored entrepreneursHIP program for the targeted girls, which provided the girls with knowledge on how to establish and run their own business. Topics included customer care, marketing, record keeping, pricing of products, personal finance, and sessions aiming at improving entrepreneurial mindset and self-confidence.

3) Cross-treatment: Information and opportunity

We also offered a cross-treatment, where the students received both the reproductive health information and entrepreneursHIP training, to investigate possible complementarities in how the two treatments work. The girls that were offered this training met twice a week for eight weeks.

3. Outcomes

We measure key outcomes in four different dimensions: *knowledge*, *behavior*, *gender-equality* and *empowerment*. Three dimensions, knowledge, behavior and gender-equality, are measured specifically for health and entrepreneursHIP, and thus we expect a stronger effect from the corresponding treatment intervention. The empowerment outcomes apply across health and entrepreneursHIP. In sum, we thus have seven outcomes in largely unrelated domains.

The domains and outcome measures are summarized in the following table and detailed below:

	Reproductive Health	EntrepreneursHIP
Knowledge	K1 *knowledge in reproductive health*	K2 *knowledge related to business practices*
Behavior	B1 *indicator of safe sex practices*	B2 *current plans to open a business*
Gender-equality	G1 *acceptance of gender-based violence*	G2 *acceptance of wife's higher earnings*
Empowerment	E1 *willingness to compete*	
	E2 *empowerment index*	

Outcomes – explanations

<u>Knowledge:</u>

K1 = knowledge in reproductive health.
We ask 7 incentivized questions about reproductive health (part 8 of the short-term follow-up questionnaire). In each question, the respondent must choose one among four possible answers. The outcome K1 is equal to the number of correct answers on these questions. Here and elsewhere we will also consider reporting the standardized outcome measure

K2 = knowledge related to business practices
We ask 5 incentivized questions about business (part 7 of the short-term follow-up questionnaire). In each question, the respondent must choose one among four possible answers. The outcome K2 is equal to the number of correct answers on these questions.

<u>Behavior:</u>

B1 = indicator of safe sex practices

This binary variable is equal to one if the respondent reports not having sex or using a condom when she has sex (which means reporting 5 or 6 in question 6.3.1 in the short-term follow-up questionnaire).

B2 = current plans to open a business
This binary variable is equal to one if the respondent reports already having made plans to open a business once the school year is completed (question 4.3 in the short-term follow-up questionnaire).

Gender equality

G1 = acceptance of gender-based violence
We ask the following question that comes from the DHS - Tanzania: "do you agree that a husband is justified in hitting or beating his wife if *(answer YES or NO)* (question 6.3.5 in the short-term follow-up questionnaire):

1) she burns the food
2) she argues with him
3) she goes out without telling him
4) she neglects the children
5) she refuses to have sexual intercourse with him"

G1 is constructed as in the DHS reports and is equal to the number of YES answers given by the respondent.

G2 = acceptance of women's higher earnings
We ask whether they agree that "it is acceptable to me that a wife earns more money than her husband" (question 5.2.6 in the short-term follow-up questionnaire). The response is on a scale of 1 to 5, where 1 is strongly disagrees and 5 is "strongly agrees". G2 is equal to the response value. This is question 5.2.6 in the short-term follow-up questionnaire.

Empowerment

E1 = willingness to compete
In an incentivized lab setting, E1 is a binary variable equal to one if the respondent chooses a competitive game rather than a fixed-payment game (question 9.2.4 in the short-term follow-up questionnaire).

E2 = Empowerment index
This index is constructed from the seven questions in 2.1 and 2.2 in the short-term follow-up questionnaire. Questions 2.2.1, 2.2.2 and 2.2.4 are inverted so that in all questions, a lower score reflects a higher empowerment.
To examine the overall impact of the interventions on the empowerment index and to account for multiple hypotheses testing, we estimate the overall average treatment effect on the index. The overall average treatment effect is estimated by combining the effects on each of questions using the method of Kling and Liebman (2004) and Kling, Liebman and Katz (2007). We will also report the estimates on the individual questions.

4. Design

We sampled schools with at least 20 girls in Form IV in the following regions: Tabora, Singida, Morogoro and Dodoma. In each school, we did a baseline survey of the girls enrolled in Form IV and surveyed the headmaster of each school, who provided us with detailed information about school characteristics.

The schools were then randomly allocated to the control group or one of the three treatments. The randomization was blocked by school-size (below or above 40 girls in Form IV) and by region.

After the baseline survey, one or two teachers per (treated) school attended a one week instructor session organized by Femina HIP (two weeks for the teachers involved in the combined treatment). After this instructor sessions, the teachers implemented the training sessions (treatments) with all the Form IV girls of their school. The single treatments had 8 training sessions of 1.5 to 2 hours, 1 session per week. The cross-treatments had 16 training sessions of 1.5 to 2 hours, two sessions per week.

The short-term follow-up survey was done within six weeks after the last training session. The survey data was collected by the girls filling in a questionnaire in the classroom. In addition, we collected incentivized data on knowledge (i.e. the students received a payment for each correct answers) and their willingness to compete.

We also surveyed 5 boys in Form IV in each school to measure spill-over effects of the treatments.

For more details, see the attached questionnaires and instructions.

5. Randomization method

We randomized at the school-level, with blocking to ensure balance in two dimensions:

- school size (more or less than 40 girls in Form IV)
- region (Tabora, Singida, Morogoro and Dodoma)

We followed David McKenzie and Miriam Bruhn's recommendations in dealing with the uneven numbers in some strata and in doing the randomization used the stata code they shared on the World Bank's "Development Impact" blog on the 11th of June 2011.

Randomization unit

We randomized at the school level. In each school, the unit of observation is all the girls enrolled in Form IV in the school-year of 2013.

Is the treatment clustered – YES/NO

YES.

6.1 Sample size – planned number of clusters

As per our contract with the funding agency, The Research Council of Norway, we planned to survey 62 schools. During the baseline, we actually surveyed 80 schools.

6.2 Sample size – planned number of observations

As per our contract with the funding agency, The Research Council of Norway, we planned to survey an average of 50 students per school. Thus a total of 3 100 students. With the increase in the number of schools surveyed, we reached a total of 3 485 students at baseline.

6.3 Sample size – number of clusters per arm

We planned to have 17 schools in the control group and 15 schools in each treatment. Having reached 18 more schools than initially planned, we randomly allocated 20 schools in each arm.

6.4 Sample size – MDE for main outcome

The sample size was powered to detect changes in pregnancy rate, which is the most demanding variable to measure and therefore serves as a conservative estimate for the other variables of interest. Pregnancy is not an outcome measured in the short-term follow up described in this pre-analysis plan, but will be covered in the planned long-term survey.

We do not presently have pregnancy rates broken down by school (which is the relevant cluster unit), but data from the DHS 2010 combined with our qualitative insights from discussions with local partners indicate that pregnancy rates among past students from Form IV should be around 15% and 35% one year after the completion of Form IV. Taking into account the effect of clustering and the fact that we have three different treatment groups in addition to a control group, we have with the planned sample a power of 80% (with a 5% confidence interval) to detect a decrease in pregnancy rate from 25% to 20% (using the approach of Hayes and Moulton, 2009).

7. Heterogeneous effects (variables from the baseline survey)

In addition to measuring the global impact of the treatments, we will also study heterogeneous effects along variables covering what we consider important dimensions in determining the impact of the intervention; the school environment, family background, and individual characteristics:

1) HET1 - School environment: remoteness, measured by a binary variable equal to 1 if it takes at least X minutes by car to reach the school from the local district headquarters, where X is the median of the distribution (and given by question 6 in the headmaster questionnaire). The binary indicator therefore identifies the most remote 50% of the schools.
2) HET2 - Family background: an index of family wealth based on
    - whether the household owns a TV (question 1.4.1 of the baseline questionnaire)
    - how many days per week do they eat meat at home (question 1.4.6 of the baseline questionnaire)
    - whether the household is connected to electricity (based on question 1.4.10 of the baseline questionnaire)

    The index is constructed by taking the average of the standardized variables on these three dimensions, where we then use a dummy for whether this index takes a value above or below the median.

3) HET3 - Individual characteristics – cognitive ability:
   We first calculate an index equal to the number of correct answers to the following questions:
   - How many zeros do you have to include if you write "twenty five million" in figures?
   - The full price of a coat is 250,000 TSH, but in a sale, the price is reduced by 20%. How much do you have to pay for the coat? ____________________ TSH
   - Write the following in order of size, starting with the smallest: 2/3 ; 65% ; 0.6

   (questions 2.1.1 to 2.1.3 of the baseline questionnaire)
   We then use a binary variable equal to one for values of the cognitive ability index above the median.

4) HET4 – Individual characteristics - age:
We here use a binary variable equal to one for values of age above the median. (question 1.1.2 of the baseline questionnaire / question 1.1.1 of the short-term follow-up questionnaire for the girls who did not attend the baseline survey)

8. Covariates (variables from the baseline survey)

Our covariates can be classified into four categories: (i) applied to all outcomes, (ii) applied to the business knowledge outcome, (iii) applied to the health knowledge outcome and (iv) applied to the self-reported behavioral outcomes (B1, B2) and the behavioral empowerment outcome (E1) measured in the lab.

(i) The following covariates will be used with all outcomes:
- HET1, HET2, HET3 and HET4 (see "heterogeneous effects" above).
- Number of girls in Form IV (question 20 in the headmaster questionnaire).
- Whether the household head is a woman (question 1.3.2. in the baseline questionnaire).
- Whether the household head owns the business (question 1.3.3 in the baseline questionnaire).

(ii) The following covariate will be used only with the health knowledge outcome (K1): Health knowledge index, equal to the number of correct answers given to four questions (part 4 in the baseline questionnaire).

(iii) The following covariate will be used only with the business knowledge outcome (K2): Business knowledge index, equal to the number of correct answers given to three questions (part 3 in the baseline questionnaire).

(iv) The following covariate will be used with the self-reported behavioral outcomes, B1 and B2, and with the behavioral empowerment outcome E1: Risk aversion, a binary indicator equal to one if the answer to question 2.2.3. in the baseline questionnaire is "I would keep 100 000 …" or "I would keep 75 000 …".

The covariates will also be used to check the balance of the short-term follow-up data over the different arms.

9. Treatment effect equation to be estimated

In the analysis, we will consider two different samples. Sample 1 consists of the students who we surveyed at baseline and the short-term follow-up. Sample 2 consists of all the students surveyed at the short-term follow-up, that is those from the baseline plus students that were missing at the baseline but attended the short-term follow up. The following equations will be estimated on both samples. We will use ordinary least squares estimators.[3]

We first regress the outcome of interest on treatment status; i.e. we include three dummies, one for each treatment group (Health, EntrepreneursHIP, and Cross treatment):

(1) $$Y_i = \alpha + \beta_1 E_i + \beta_2 H_i + \beta_3 CROSS_i + \varepsilon_i$$

We cluster the standard errors at the school-level. Since we do not expect any negative treatment effects, we will use one-sided tests of the treatment coefficients in our main specification.

When testing whether the impact of the cross treatment is equal to the two separate treatments, we do not have a clear a priori hypothesis about the sign, and thus we will use a two-sided test to test whether
$\beta_1 + \beta_2 = \beta_3$.

Furthermore, we will also estimate equation (1) with a set of covariates **X:**

(2) $$Y_i = \alpha + \beta_1 E_i + \beta_2 H_i + \beta_3 CROSS_i + \beta_4 \boldsymbol{X}_i + \varepsilon_i$$

The covariates are detailed above.

In all estimations, if we have the lagged dependent variable, we will include it among the covariates.
Finally, we will also study heterogeneity in treatment effects using the variables defined in the "heterogeneous effects" section above. We will then introduce interaction terms, where the three treatment arms will be interacted with the relevant variable. When checking for heterogeneous effects, the equation becomes:

(3) $$Y_i = \alpha + \beta_1 E_i + \beta_2 H_i + \beta_3 CROSS_i + \beta_4 X_i + \beta_5 W_i + \beta_6 W_i * E_i + \beta_7 W_i * H_i + \beta_8 W_i * CROSS_i + \varepsilon_i$$

Where W stands for the variable defining the heterogeneous effects of interest. Equation (3) will also be estimated without the covariates X.

We will run both separate regressions for each of the background variables, where only the interaction terms for this background variable is added to (2), and a joint regression

[3] Note that some of the covariates and heterogeneous effects variables defined above are not available for both sample.

including interaction terms for all the background variables introduced in the "heterogeneous effects" section above

10. Dealing with multiple outcomes

Our seven key outcomes are defined on largely unrelated domains and thus we do not adjust for multiple inference across the domains, as we view them as conceptually distinct. . Within each domain, with the exception of E2, we only focus on a single outcome variable. We therefore only make correction for multiple outcomes for E2.

When analyzing outcome E2, we follow the approach of Kling, Liebman and Katz (2007) and Kling and Liebman (2004), and create standardized treatment effects within the domain. Specifically, we follow the exact procedures of Finkelstein et al (2012). When reporting individual p-values within the domain, we report both the unadjusted and the family-wise adjusted p-values.

11. Addressing survey attrition and non-response

We will check whether survey attrition is correlated with the treatments. If that is the case, we will estimate lee-bounds (Lee 2009).[4] We will also follow Kling, Liebman and Katz (2007). We obtain lower bounds of the treatment effects by replacing missing observations in the treatment (control) arms by the corresponding arm's mean value minus (plus) 0.05, 0.10 and 0.20 standard deviations of the control group. Upper bounds of the treatment effects are constructed in a symmetrical way.

No imputation for missing data from item non-response at follow-up will be performed. We will check whether item non-response is correlated with treatment status following the same procedures as for survey attrition, and if it is, construct bounds for our treatment estimates that are robust to this.

12. Dealing with outcomes with limited variation

We follow David McKenzie's approach: "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. In the event that omission decisions result in the exclusion of all constituent variables for an indicator, the indicator will be not be calculated" (Development Impact blog, The World Bank, October 28th 2012).

13. Spill-overs to boys

To measure spill-overs to boys of the treated school, we surveyed five boys per school in the short-term follow-up survey. The boys were selected by the teacher to whom we instructed to: "Please, ask five boys to be present also. We will ask them to do a small math test." That math test is used in the lab game played by the girls, see detailed instructions. In addition, we asked the boys to answer the questions that define outcomes G1 and G2.

[4] In the short-term follow-up, we put great effort into limiting attrition and we revisited the schools (mainly from the control group) that had a lower attendance during the survey.

Equations (1) and (2), and the corresponding heterogeneity regressions, will also be estimated on the sample of boys, where outcomes G1 and G2 will be the dependent variables. We will here have to restrict ourselves to the school characteristics covariates, since we did not collect individual background information on the boys.

14. ITT – ATE

We focus on intention-to-treat effects in our key results. Indeed, we could not force all the students to participate to each and every session, and in the exploratory analysis we will also discuss ATE-effects.

15. OTHER EXPLORATORY OUTCOMES

Our short-term follow up survey measures many other outcomes that we will use in the exploratory analysis, to complement our main findings.

In particular, we will look in more details into the competition game played in a lab-setting by all the students. While the decision to compete is one of our key outcomes, we will in addition check whether the treatments affected the girls' ability to take the "correct" decision in a competitive setting, following the approach of Niederle and Vesterlund (2007).

16. REFERENCES

Anderson, Michael, ''Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects,'' Journal of the American Statistical Association, 103, 1481–1495, 2008.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, "The Oregon Health Insurance Experiment: Evidence From The First Year", The Quarterly Journal of Economics, 127 (3), p.1057-1106, 2012.

Hayes, Richard and R.J. Moulton, "Cluster Randomized Trials", Chapman & Hall, 2009.

Kling, Jeffrey, and Jeffrey Liebman, ''Experimental Analysis of Neighborhood Effects on Youth,'' Mimeo, available online at http://www.nber.org/mtopublic/483.pdf, 2004.

Kling, Jeffrey, Jeffrey Liebman, and Lawrence Katz, ''Experimental Analysis of Neighborhood Effects,'' Econometrica, 75, 83–119, 2007.

Lee, David, "Training, wages, and sample selection: Estimating sharp bounds on treatment," NBER Working Paper No. 11721, 2009.

Niederle, Muriel and Lise Vesterlund, "Do Women Shy Away From Competition? Do Men Compete Too Much*?",* The Quarterly Journal of Economics*, 122 (3), p.1067-1101,* 2007.

Westfall, Peter H., and Stanley S. Young Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment, New York, Wiley, 1993.