Educate Girls Development Impact Bond Evaluation
Pre-Analysis Plan

September 2015

Version 3: This document was drafted during data collection for the baseline survey.

Please direct any questions about this document to Jeff McManus, Senior Technical Manager, jeffery.mcmanus@idinsight.org

Contents

I. Evaluation objectives

Primary objective: Estimate learning gains attributable to Educate Girls

The primary objective of the evaluation is to estimate the causal impact of the Educate Girls (EG) program on aggregate learning gains over a three-year period among students in grades 3-5 in government primary schools. Our impact estimates will be used to determine fair outcome payments for the Educate Girls Development Impact Bond (DIB)[1]. We will conduct a clustered randomized controlled trial (RCT) to estimate learning gains caused by the EG program.

A caveat to our estimate of learning gains

[1] See the official press release Instiglio's website for more information about the DIB.

We are likely to slightly overestimate the impact of EG on learning gains, and the DIB Working Group[2] has agreed that this is necessary in order to incentivize EG to target services to the highest-need children and to make fair payments on the DIB.

Our evaluation will yield unbiased estimates of learning gains attributable to EG for students who are enrolled at baseline. However, EG's program includes efforts to enroll out-of-school girls (OOSGs), and so we expect more newly-enrolled students to join treatment schools than control schools over the course of the evaluation. These newly-enrolled students will likely bring the average learning levels of treatment schools down at endline; by including these students in a simple difference-in-differences estimate of learning gains, we would underestimate the impact of EG and incentivize EG to keep unenrolled children out of school, but by excluding these students in our estimates, we would fail to credit EG with their learning gains.

The DIB Working Group has therefore decided to keep newly-enrolled students in the calculation of learning gains and to make a strong assumption about their potential outcomes: that in the absence of the EG program, these students would have received the lowest possible score on student assessments. Although this assumption will likely lead to an overestimate of the impact of the EG program among newly-enrolled students (and thus a slight overestimate of the impact of the EG program on all students), the DIB Working Group has agreed with this approach since it is an implicit incentive for EG to enroll this priority demographic.

Another implication of differential enrollment rates across evaluation groups is that treatment schools will likely have more students than control schools at endline. In order to credit EG with the learning gains of all students, we will therefore calculate an aggregate treatment effect (i.e. the sum of the changes in learning levels for all students) rather than an average treatment effect.

Finally, we will estimate the difference in learning gains using the difference-in-differences estimator. Although ANCOVA is likely to be more efficient, the DIB Working Group agreed on using the difference-in-differences estimator since it is easier to interpret for determining outcome payments.

Secondary objective: Estimate the change in enrollment of girls in treatment villages

We will also estimate the change in enrollment of OOSGs in treatment villages. Due to budgetary and logistical constraints, which prevent us from conducting an OOSG census in control villages, we will only measure enrollment before and after the evaluation in treatment villages. Since there may be other factors besides the EG program that explain changes in enrollment in treatment villages, we cannot fully attribute changes in enrollment to the causal impact of the EG program.

Enrollment outcomes will determine approximately 20% of DIB payments, whereas learning outcomes will determine approximately 80% of DIB payments. Since we are using a non-experimental method to measure enrollment outcomes, which in any case play a relatively minor role in the DIB, the remainder of this pre-analysis plan will focus on learning outcomes. Details on our measurement of enrollment outcomes are available upon request.

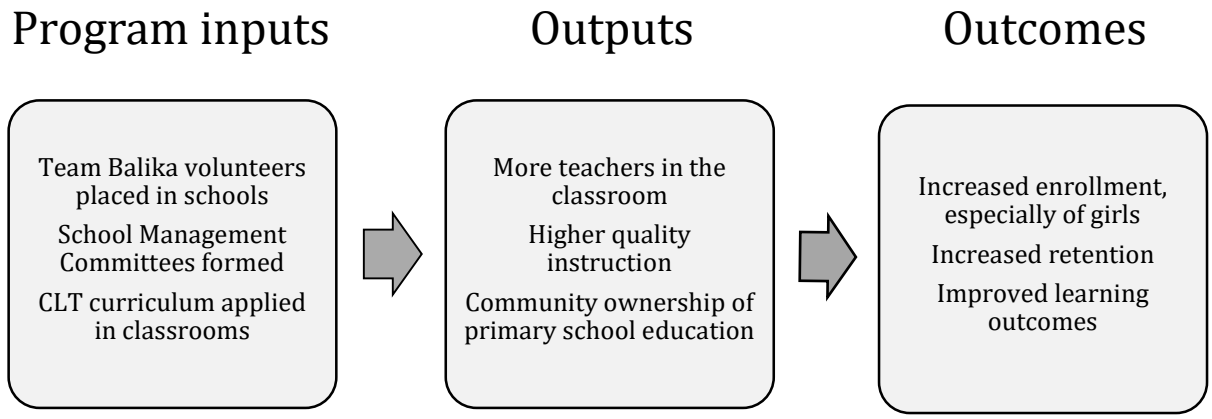
[2] The DIB Working Group includes Instiglio (the Project Manager), the Children's Investment Fund Foundation (the Outcome Payer), the UBS Optimus Foundation (the Investor), Educate Girls (the Service Provider), and IDinsight (the Evaluator).

II. The Educate Girls program

EG is an eight-year-old nonprofit organization that focuses on enrollment, retention and learning of marginalized girls by leveraging existing government and community infrastructure. EG is currently present in six critical gender gap districts in Rajasthan, and its staff operate in 4,500 villages and 8,000 schools. We will evaluate the EG program as it is rolled out to a new district in Rajasthan, Bhilwara district, in 2015.

The core of EG's intervention is the Creative Learning and Teaching (CLT) program. The CLT program aims to improve learning outcomes among boys and girls in grades 3-5 in government primary schools. Developed in collaboration with Pratham, the CLT program employs two key pedagogies: group work and closing the learning gap by allocating extra resources to the lowest-performing students. EG will train volunteers ("Team Balika") to deliver the CLT program three times per week in treatment schools. **Figure 1** shows a simplified Theory of Change of the EG program on learning gains:

Figure 1: Theory of Change for EG's program



III. Evaluation methodology

Sampling and randomization

The impact of EG's program on learning gains will be estimated using a clustered RCT, with villages randomly assigned to either the treatment group (all eligible schools within a treatment village receive the EG program) or the control group (all eligible schools within a control village do not receive the EG program). Since several EG activities take place at the village-level, the randomization has occurred at the village level rather than at the school level or individual level. The following process summarizes how villages and students will be sampled and randomized into treatment and control groups:

Activities already completed

1. **Prepare the sampling frame**

a. ***Select the geographic region***: Bijoliya, Jahajpur, and Mandalgarh blocks in Bhilwara district, Rajasthan. This region was selected based on its similarity with areas where EG has worked previously in terms of low educational outcomes and a pronounced gender gap, and based on its similarity with areas where EG is likely to expand in the future.

b. ***Filter out ineligible villages*** according to the following criteria:

- Village eligibility criteria (based on the 2014-15 DISE database[3]):
  - Has at least 1 eligible school (see below) and no more than 4 eligible schools. Villages with more than 4 eligible schools are generally too large for EG to effectively apply their program.
  - In a rural area (where all schools in the village are listed as 'rural' in the DISE database). EG's program is not applicable to urban areas.

- School eligibility criteria (based on the 2014-15 DISE database):
  - Includes grades 1-5 (Includes schools that are 'primary only' and 'primary with upper primary'. Excludes schools that are 'primary with upper primary and secondary' and 'primary with upper primary and secondary and higher secondary'). EG's program is targeted at primary-school students.
  - Under the management of the Department of Education or Local Bodies. (Excludes private schools, madrasas, Sanskrit schools and Shisksha Karmi schools). EG's program must be conducted in government-run schools.
  - Has at least 10 and no more than 60 primary enrolled students in grades 3-5, according to government records. Schools with fewer than 10 students are too small and schools with more than 60 students are too large for EG to effectively apply their program.
  - Medium of instruction is Hindi. All of EG's materials are in Hindi.
  - Operationally feasible (excludes schools that EG says are not possible to deliver services to).

- Other criteria:
  - The school exists in the 2013-14 DISE database. Schools that do not exist in the 2013-14 database are very new and may not have the infrastructure necessary to support the EG team.
  - The school is not listed as 'urban' in the 2013-14 DISE database. Schools in areas that were previously listed as 'urban' are likely in peri-urban regions that would be difficult for EG to work in.

c. ***The remaining villages and schools comprise the sample***.

2. Assign treatment status

[3] The District Information System for Education (DISE) collects and posts data annually on all government schools in India. Download the data on their website: http://schoolreportcards.in

a. ***Create strata*** defined by block and number of schools in each village. For strata with an odd number of villages, randomly select one village to be dropped.

b. ***Match pairs of similar villages within strata*** based on scores from the first component of a principal components analysis on the following attributes: (i) total enrollment of primary school students, (ii) the presence of an upper primary school, (iii) percent of enrolled primary school students who are female, and (iv) student-to-teacher ratio. These attributes were selected based on our understanding of the likely drivers of educational outcomes in this setting. We randomize using pairwise matching because it leads to more balanced evaluation groups than simple random assignment.

c. ***Randomly assign one village in each matched pair in the sample to treatment***.

d. ***Conduct t-tests on matching variables to check that treatment and control groups are balanced.*** If an unexpectedly large number of t-tests (>20%) find significant differences at the 5% level between treatment and control groups, then re-randomize. (The differences between treatment and control groups for all village-level and school-level attributes were not statistically significant.)

3. ***Drop village pairs based on budgetary constraints***. Steps 1 and 2 yielded a sample of 396 schools in 338 villages. Due to budgetary constraints, the evaluation sample had to be reduced to 332 schools in 280 villages. 28 village-pairs were randomly selected to be dropped to meet the reduced sample size requirements.

Activities currently in progress

4. ***Sample students within selected schools***. Randomly sample 50% of boys and 50% of girls in each grade for grades 1-5 from all eligible schools in sampled villages. If there are fewer than 4 boys/girls in a grade, then sample all boys/girls in that grade. Administer baseline assessments to these students.

Activities to be completed

5. ***Administer endline assessment to all students sampled at endline who are present, and administer assessments at home to a sample of students who are absent at endline***. We will administer assessments at home to a random subset of students to ensure that differential attrition does not introduce bias. The precise number of home assessments will be determined based on enumerator availability and budget constraints.

This design yields a minimum detectable effect size of approximately 0.5 ASER points (0.2 SD). Full power calculations are available upon request.

Validity of estimates generated from this process

Steps 1-3 in this process generated experimental groups that are roughly balanced along observable characteristics that are likely to affect outcomes (**Table 1**):

Table 1: Balance checks on sample

Variable	Control Mean	Treatment Mean	p_value of Difference
Village-level variables			
ENROLLMENT GRADES 1-5	54.07	53.39	0.84
ENROLLMENT GRADES 1-5 GIRLS (% OF TOTAL)	0.51	0.51	0.83
VILLAGE HAS AN UPPER PRIMARY SCHOOL	0.56	0.56	0.90
STUDENT-TEACHER RATIO	19.79	19.29	0.74
School-level variables			
ENROLLMENT GRADES 1-5	45.60	45.02	0.80
ENROLLMENT GRADES 1-5 GIRLS (% OF TOTAL)	0.52	0.51	0.78
INCLUDES UPPER PRIMARY GRADES	0.38	0.39	0.82
STUDENT-TEACHER RATIO	20.17	19.59	0.68
MANAGED BY DEPARMENT OF EDUCATION	0.38	0.39	0.91

Based on this evidence, and the assumption that the experimental groups are also roughly balanced on key unobservables, we will likely be able to obtain unbiased estimates of learning gains attributable to EG for students enrolled *at baseline*. However, a core EG village-level activity is enrollment of OOSGs, and so we anticipate large changes in class compositions in treatment villages over the course of the study.[4] Because of this, enrolled students at endline in treatment and control villages will not be comparable to each other: Since newly-enrolled students will not have baseline data and would likely score lower than the average originally-enrolled student, a simple comparison of treatment and control averages at endline would likely underestimate the impact of the EG program. In order to estimate aggregating learning gains to determine outcome payments for the DIB, we will therefore estimate learning gains separately for originally-enrolled and newly-enrolled students (the latter estimates will require additional assumptions), and sum them. See below for more details.

IV. Data collection

Learning gains will be measured using the Annual Status of Education Report (ASER) assessment tool. The ASER assessment consists of three sections: Hindi, Math, and English. Each section consists of 4 levels (and a possible score of 1 to 5 points); we have added one extra level to the Hindi section to reduce the frequency of "ceiling effects", in which many students obtain the highest possible score on the section, leading to underestimates of learning gains.[5] Students can therefore earn between 3 and 16 points on the assessment. **Table 2** shows how the ASER test will be scored for a hypothetical student:

[4] In previous EG districts, by the end of one year of implementation, newly enrolled girls expanded school rosters by 20-30%.

[5] A separate study that used the ASER assessment to measure learning gains among students in government schools in Rajasthan found that approximately 40% of students in Standard 5 obtained the highest possible score on the Hindi section, suggesting that the assessment may not capture higher-level outcomes for some students. There did not appear to be similar levels of censoring on the English and Math sections.

Table 2: Example Student #1234, Endline ASER Test Score

Section	Competencies (highest level attained in bold)	Points earned
Hindi	Beginner Letter recognition **Word recognition** Paragraph fluency Story fluency Story Plus fluency	3
Math	Beginner Numbers 0-9 recognition Numbers 10-99 recognition **Subtraction** Division	4
English	**Beginner** Capital letter recognition Small letter recognition Word recognition Sentence fluency	1
Total score		***8***

Our enumerators are following the instructions on the ASER website (found here) to administer the assessment.

In order to improve the precision of our estimates of the impact of the EG program, we are conducting baseline assessments for all originally-enrolled sampled students. Since students will be leaving the program over the course of the evaluation (e.g. students in Grade 5 will exit after 1 year), we will conduct one endline assessment at the end of each year. The following steps summarize the data collection process:

1. During the Year 1 baseline, currently in progress, our enumerators are visiting all treatment and control schools and administering an ASER assessment to a randomly-selected sample of 50% of boys and girls in each grade for grades 1-5 (or to all students if there are fewer than 4 boys/girls in a grade). Our teams are visiting both treatment and control schools in the same geographic cluster on the same day in order to ensure similar testing conditions across evaluation groups. Students will receive a grade (out of 5 or 6 points) on each of the three sections of the assessment (English, Hindi, and Math), and their overall grade will be the sum of these grades (possible scores: 3-16).

2. During the Year 1 endline, we will administer assessments to all previously-sampled students in grades 3-5 in treatment and control schools, and also for 50% of newly-enrolled students.

3. Some sampled students who are enrolled in grade 5 may be absent on the day of the Year 1 endline. In order to obtain an unbiased estimate of the learning gains of these students, we will take a random sample of these students (to be determined based on budget constraints

and the fraction of students with missing scores) and administer the assessment at their homes. The same fraction of missing students will be sampled in each school.

4. Steps (2) and (3) will be repeated during the Year 2 and Year 3 endlines. During the Year 3 endline, sampled students enrolled in grades 3 and 4 who were absent for the assessment will also be sampled in order to obtain unbiased estimates of their learning gains (the same fraction of students with missing endline scores in each grade in each school will be sampled).

Table 3 shows data collection activities within the broader context of the evaluation:

Table 3: EG DIB evaluation activities

		2015				2016				2017				2018	
Tasks	**Deadlines**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**
Evaluation design															
Treatment assignment	Feb 2015	■													
Finalize evaluation design	Mar 2015	■													
Prepare data collection instruments	Jul 2015		■	■											
Field work															
Y1 baseline survey	Sept 2015			■											
Y1 endline survey	Jan 2016					■									
Y2 endline survey	Jan 2017									■					
Y3 endline survey	Jan 2018													■	
Analysis and reporting															
Y1 interim report	Mar 2016						■								
Y2 interim report	Mar 2017										■				
Final evaluation report	Mar 2018													■	
Final presentation of results	Apr 2018														■

V. Analytical approach

Evaluation objective: *Estimate the causal impact of the EG program on aggregate learning gains over a three-year period among students who are enrolled in grades 3-5 at any point during the evaluation.*

Unit of analysis

Student (all students who are enrolled in grades 3-5 at any point during the evaluation)

Outcome measurement

Learning gains, as measured by the change in scores on repeated ASER assessments

Estimation strategy

Estimates of the aggregate treatment effect will be calculated using a difference-in-differences estimator. We will calculate learning outcomes separately for students depending on whether they are present or absent at baseline and endline, resulting in five different *types* of students:

- Type 1 students: Enrolled in grades 1-5 at baseline, present for baseline, present at endline
- Type 2 students: Enrolled in grades 1-5 at baseline, present for baseline, enrolled but absent at endline
- Type 3 students: Enrolled in grades 1-5 at baseline, present for baseline, drops out early or is inaccessible at endline (e.g. migrates out of village)
- Type 4 students: Absent or unenrolled at baseline, present at endline
- Type 5 students: Absent or unenrolled at baseline, enrolled but absent at endline

Learning outcomes for each of these types of students will be calculated as follows:

Type I students

Enrolled in grades 1-5 at baseline, present for baseline, present at endline

Estimation strategy

$Learning\ gains\ of\ Type\ I\ students = \sum_{i_t}^{I_t}(S_{E,i} - S_{B,i}) \ - \ \sum_{i_c}^{I_c}(S_{E,i} - S_{B,i})$, where

- $S_{E,i}$ is the applicable endline ASER score for Type I student *i*:
  - Y1 endline for students in grade 5 at baseline
  - Y2 endline for students in grade 4 at baseline
  - Y3 endline for students in grades 1-3 at baseline
- $S_{B,i}$ is the Y1 baseline ASER score for Type I student *i*
- *t* denotes students in treatment schools and *c* denotes students in control schools

Type I student learning outcomes in one hypothetical school

	Grade at baseline	Y1 Baseline	Y1 Endline	Y2 Endline	Y3 Endline	Subtotal
Student 1	1	8	-	-	10	**+2**
Student 2	2	5	-	9	11	**+6**
Student 3	3	4	6	6	5	**+1**
Student 4	4	11	12	11	-	**0**
Student 5	5	8	7	-	-	**-1**
					Learning gains	**+8**

Type II students

Enrolled in grades 1-5 at baseline, present for baseline, enrolled but absent at endline

Estimation strategy

$Learning\ gains\ of\ Type\ II\ students = \sum_{ii_t}^{II_t} W(S_{E,ii} - S_{B,ii}) \ - \ \sum_{ii_c}^{II_c} W(S_{E,ii} - S_{B,ii})$, where

- All similar variables are defined as above

- *ii* is a Type II student randomly selected to receive a follow-up survey at home and *II* is the total number of Type II students who receive follow-up surveys[6]
- *W* is the sampling weight applied according to the fraction of Type II students selected for follow-up out of all Type II students.

Type II student learning outcomes in one hypothetical school

Shown for students in Grade 3 at baseline, though the same exercise would be conducted for students in all grades at baseline.
Students highlighted in gray if they are randomly selected to be administered follow-up assessments at home. Assumes 1/3 of all missing students are selected for follow-up.

	Grade at baseline	Y1 Baseline	Y1 Endline	Y2 Endline	Y3 Endline (Re-surveys)	Subtotal
Student 1	3	8	8	11	-	**-**
Student 2	3	9	8	9	11	**+2**
Student 3	3	9	11	15	-	**-**
Student 4	3	4	6	8	7	**+3**
Student 5	3	6	5	5	-	**-**
Student 6	3	6	10	14	-	**-**
Learning gains (with sampling weights)						**(2+3)*3 = +15**

Type III students

Enrolled in grades 1-5 at baseline, present for baseline, drops out early or is inaccessible at endline (e.g. migrates out of village)

Estimation strategy

$Learning\ gains\ of\ Type\ III\ students = \sum_{iii_t}^{III_t}(S_{E_L,iii} - S_{B,iii}) - \sum_{iii_c}^{III_c}(S_{E_L,iii} - S_{B,iii})$, where

- All similar variables are defined as above
- *iii* is a Type III student and *III* is the total number of Type III students
- E_L denotes the last endline for which a Type III student's scores are available (students who drop out before grade 3 will be dropped from the study). Note that this is not necessarily the last endline for which the student is enrolled (e.g. if the student is enrolled in Y1 and Y2 and drops out after Y2, but is absent for the Y2 endline). The reason for not following up with these students is because (1) IDinsight will not be able to anticipate when they will drop-out and (2) once they drop out they may be very costly to reach (because they may migrate out of the village, etc.). Hence, the treatment effect for these students might be underestimated if a large number of them are absent for their last assessment.

Type III student learning outcomes in one hypothetical school

	Grade at	Y1	Y1	Y2	Y3	Subtotal

[6] The size of the random sub-sample will be determined based on budget considerations and the fraction of total students with missing scores. The same fraction of missing students in each grade and in each school will be selected for follow-up at home.

	baseline	Baseline	Endline	Endline	Endline	
Student 1	1	2	-	-	-	**-**
Student 2	2	6	-	8	-	**+2**
Student 3	3	8	9	9	-	**+1**
Student 4	4	5	7	-	-	**+2**
Student 5	5	13	-	-	-	**-**
					Learning gains	**+5**

Type IV students

Absent or unenrolled at baseline, present at endline

Estimation strategy

$Learning\ gains\ of\ Type\ IV\ students = \sum_{iv_t}^{IV_t}(S_{E,iv} - 3) - \sum_{iv_c}^{IV_c}(S_{E,iv} - 3)$, where

- All similar variables are defined as above
- *iv* is a Type IV student and *IV* is the total number of Type IV students (in treatment or control schools)
- 3 points are subtracted from all scores since it is the lowest possible score on the assessment.
- Note that this calculation will likely overestimate the effect of treatment for Type IV students since it includes students who enroll as a consequence of treatment (and the corresponding children in the control group will be unenrolled and not identifiable, and therefore untested, and so the implied assumption is that they would have obtained the lowest possible score on the assessment). This overestimation can be viewed as giving greater weight to the outcomes of newly-enrolled students than the outcomes of other students; since these students are of special interest in EG's theory of change, the Working Group has agreed to pay this implicit premium for their learning outcomes.

Type IV student learning outcomes in in one hypothetical school

	Grade at baseline	Y1 Baseline (imputed)	Y1 Endline	Y2 Endline	Y3 Endline	Subtotal
Student 1	1	3	-	-	5	**+2**
Student 2	2	3	-	5	7	**+4**
Student 3	3	3	4	6	7	**+4**
Student 4	4	3	6	9	-	**+6**
Student 5	5	3	9	-	-	**+6**
					Learning gains	**+22**

Type V students

Absent or unenrolled at baseline, enrolled but absent at endline

Estimation strategy

$Learning\ gains = \sum_{v_t}^{V_t} W * (S_{E,v} - 3) - \sum_{v_c}^{V_c} W * (S_{E,v} - 3)$, where

- All similar variables are defined as above
- *v* is a Type V student randomly selected receive a follow-up survey at home and *V* is the total number of Type V students who receive follow-up surveys
- 3 points are subtracted from all scores since it is the lowest possible score on the assessment.
- *W* is the sampling weight applied according to the fraction of Type V students selected for follow-up out of all Type V students.
- As with Type IV students, this calculation will likely overestimate the effect of treatment for Type IV students since it includes students who enroll as a consequence of treatment (and the corresponding children in the control group will be unenrolled and not identifiable, and therefore untested, and so the implied assumption is that they would have obtained the lowest possible score on the assessment). This overestimation can be viewed as giving greater weight to the outcomes of newly-enrolled students than the outcomes of other students; since these students are of special interest in EG's theory of change, the Working Group has agreed to pay this implicit premium for their learning outcomes.

Type V student learning outcomes in one hypothetical school

Shown for students in Grade 3 at baseline, though the same exercise would be conducted for students in all grades at baseline.

Students highlighted in gray if they are randomly selected to be administered follow-up assessments at home. Assumes 1/3 of all missing students are selected for follow-up.

	Grade at baseline	Y1 Baseline (imputed)	Y1 Endline	Y2 Endline	Y3 Endline (Re-surveys)	Subtotal
Student 1	3	3	5	8	-	**-**
Student 2	3	3	-	4	6	**+3**
Student 3	3	3	-	-	-	**-**
					Learning gains	**+3*3 = +9**

The aggregate treatment effect is the sum of all types of students' learning gains.