Request for Proposals for the Evaluation of Educate Girls’ Five-Year Scale-Up in India

Deadline for questions from interested organization: October 31, 2019 at 23:59 PDT
Proposal deadline: November 18, 2019 at 23:59 PDT
All questions and proposals should be sent to jeffery.mcmanus@idinsight.org

I. OVERVIEW

IDinsight seeks to fund an evaluator to undertake an impact evaluation and collect verification data on the five-year scale-up of Educate Girls’ program in India. Educate Girls is a Mumbai-based non-profit that helps to enroll girls into government primary and upper primary schools and delivers a remedial tutoring curriculum through community-based volunteers. Educate Girls was selected as a 2019 Audacious Project recipient to enable the program to scale from the current 13,000 villages to approximately 35,000 villages in pursuit of the following targets:

1. Enrollment of 1.56 million out-of-school girls into grades 1 through 10
2. Retention of 1.33 million newly-enrolled girls
3. Improvement in learning outcomes for 900,000 girls and boys in grades 3 through 10
4. Development of 77,000 adolescent girls with Life Skills Education
5. School governance and infrastructure improvement in 40,000 schools that serve 15.8 million boys and girls

These Terms of Reference describe an evaluation agenda to assess whether Educate Girls successfully delivers on these targets. The evidence from these evaluation activities is critical to track progress, inform course correction, and maximize the effectiveness of this ambitious scale-up. We seek an evaluator to conduct the following activities:

1. An impact evaluation to estimate the causal effect of Educate Girls’ scale-up on learning outcomes in a representative sample of villages that receive the primary learning program (Target 3)
2. Verification data collection to validate Educate Girls’ monitoring data on enrollment and retention in a representative sample of all program villages (Targets 1 and 2)
3. Verification data collection to validate Educate Girls’ monitoring data on implementation activities in a representative sample of villages that receive life skills programming (Target 4)
4. Verification data collection to validate the submission of school improvement plans in a representative sample of villages that are included in the school governance and infrastructure improvement programming (Target 5)
II. BACKGROUND

1. EDUCATE GIRLS PROGRAM

Educating girls advances many development goals: Educated women are healthier, earn higher incomes, marry later, have fewer children, and invest more in the health and education of their children.¹ Yet in India, despite recent strides to improve enrollment, 4.1 million girls remain out of school, and girls are more likely than boys to be non-enrolled. Even when girls (and boys) show up for school, they often fail to master even basic reading and math skills. According to the 2018 Annual Status of Education Report (ASER), which assesses a representative sample of students in every district in India, only 27% of children in grade 3 and 50% of children in grade 5 can read at the grade 2-level. 72% of children in grade 5 are unable to complete a subtraction problem from the grade 2 curriculum.

Educate Girls seeks to address these educational inequities by recruiting and managing a large staff of field coordinators and a network of community-based volunteers to deliver their program to tens of thousands of the poorest and most remote villages in India. When Educate Girls expands to a new village, field coordinators recruit a “Team Balika” (or “Team for Girls”) community volunteer from among the most educated young men and women in the village. Team Balika volunteers receive intensive training from Educate Girls staff on family counselling on the value of education, community mobilization techniques, and on the specific curriculum that they will deliver in government schools. After training, Educate Girls staff, supported by Team Balika volunteers, conduct a door-to-door census of the village to identify all out-of-school girls.

After a year of preparation activities (referred to as “Year 0” in this document), the Team Balika volunteer and Educate Girls staff conduct several activities in the village to support the educational outcomes of out-of-school girls and children in school, including:

- Counselling with families that have out-of-school girls age 5 to 14 to encourage and support enrollment into government primary and upper primary schools.
- Community mobilization events to inform the community on the socioeconomic benefits of girls education and shift the community mindset on the importance of girls’ education.
- Delivery of a basic reading and math curriculum to students in grades 3 to 5 in a subset of government primary schools in enrollment villages. Educate Girls’ curriculum, “Gyan ka Pitara” (or “Repository of Knowledge”), is intended to supplement the state curriculum and help students who lag behind (including newly-enrolled girls) catch up to their peers. The curriculum is based around worksheets that volunteers assign to students according to individual need.
- Creation of “Bal Sabhas” or “Girl Leadership Councils” for girls in Grades 6, 7, and 8. Girls elected to these 13-member councils receive “life skills education” from volunteers to boost confidence, communication, leadership, public speaking, decision-making, and problem-solving skills.

---

• Supporting “School Management Committees” (essentially Parent Teacher Associations) to improve school governance and facilities, with an emphasis on the preparation and submission of School Improvement Plans to build facilities that are critical for improving attendance and retention of newly enrolled girls, such as girls’ bathrooms, drinking water facilities, and kitchen sheds.

Starting in 2019, Educate Girls began piloting a new program component, the Adolescent Girls Program (AGP), in villages where they had operated their primary program in primary and upper primary schools for several years. Educate Girls staff observed that girls often dropped out of school during the critical transition from upper primary to secondary school. The AGP extends Educate Girls’ program to girls aged 10 to 19 and supports them in this transition. Educate Girls is also piloting a new life skills curriculum for the AGP, which will deliver leadership training to girls through “Kishori Samuhs” or adolescent girl leadership councils. For the purposes of this evaluation, the evaluator will only be expected to verify the delivery of life skills education in AGP, though it will not be included in the impact evaluation.

For more information on Educate Girls’ programming, see the Educate Girls website.

From 2015 to 2018, Educate Girls participated in the world’s first development bond (or “DIB”), the Educate Girls DIB. As part of this DIB, IDinsight conducted a three-year randomized controlled trial of Educate Girls program across 332 schools, which found that Educate Girls’ program had large positive effects on enrollment and learning in primary schools. The evaluation described in these Terms of Reference will assess whether Educate Girls can maintain high levels of impact at a much larger scale outside of the DIB framework, and with an updated program model. We encourage the evaluator to draw on the lessons from the DIB evaluation during design, data collection, and analysis; see Appendix 2 for more information on the DIB evaluation.

2. ANTICIPATED PROGRAM SCALE-UP

*The details of this scale-up plan are subject to change. Any changes would be discussed with the evaluator to ensure that evaluation activities remain within the agreed upon scope of work.*

Educate Girls has been selected as a 2019 Audacious Project recipient to scale their program and enrol an additional 1.56 million out-of-school girls by 2024, which will likely occur across 4 states (Rajasthan, Madhya Pradesh, Uttar Pradesh, and Bihar). This scale-up will include five program components:

1. **Existing Villages Program:** Educate Girls will continue to implement enrollment and retention activities and the primary school learning program (“GKP”) in villages where they have been operating in Rajasthan and Madhya Pradesh. Over the course of the project Educate Girls will phase out of many of these villages.

2. **Adolescent Girls Program:** As Educate Girls phases the primary program out of Existing Villages, many of these villages will begin to receive the Adolescent Girls Program, described above. Villages will be identified by Educate Girls to receive the AGP based on geography, personnel decisions, and the presence of upper primary and secondary schools. Once introduced, the AGP will continue to operate in selected villages for up to three years.
3. **Primary Organic Program**: In a subset of new villages, Educate Girls will implement enrollment and learning activities through field coordinators and Team Balika volunteers. Using machine learning algorithms developed in collaboration with IDinsight, Educate Girls will expand the Primary Organic Program to new regions each year that are predicted to have the most out-of-school girls. Within these regions, clusters of approximately 8 to 10 villages (called “Field Coordinator clusters” or “FC clusters”) will be identified based on geographic proximity and to ensure a roughly balanced case load of out-of-school girls across FCs. The enrollment, retention, and school infrastructure components of the program will be implemented in all villages in FC clusters, starting in “Year 1” of implementation (i.e. the year after EG has conducted the door-to-door survey in all villages in the FC cluster). However, the learning curriculum will be introduced in Year 1 in only two of the approximately 8 to 10 primary schools per FC cluster. An additional two primary schools per FC cluster will receive the learning curriculum in Year 2 of the program, and a final two primary schools per FC will receive the learning curriculum in Year 3 of the program. As described below, the evaluator is encouraged to leverage school selection within FC clusters to form a counterfactual.

4. **Primary Inorganic Program**: In a different subset of new villages, Educate Girls will implement enrollment and learning activities through local non-profit partners. These villages will also be identified by Educate Girls using predictive algorithms in collaboration with IDinsight. Educate Girls will provide guidance and monitoring to ensure that partner organizations implement the Primary Inorganic Program in the same way as the Primary Organic Program. Villages and schools in the Primary Inorganic Program should be included in both verification data collection and the impact evaluation.

5. **Technical Partnerships**: In a final subset of new villages, Educate Girls will collaborate with local government through technical partnerships to implement enrollment and retention activities. These villages will be identified in conjunction with the government and will not be targeted using predictive algorithms. Educate Girls will train local government staff to administer the door-to-door census to enumerate out-of-school children and to provide family counseling and community mobilization. The learning program and school governance and infrastructure support will not take place in Technical Partnership villages, and so for the purposes of this evaluation, the evaluator is expected to only verify enrollment and retention in a representative sample of these villages.

Table 1 summarizes the number of villages expected in each program subcomponent over the course of the five-year scale-up. *These numbers are intended to be illustrative of the scale of Educate Girls’ expansion via different program components, but the exact number of villages within each component is subject to change as Educate Girls continues to refine their delivery model.*

<table>
<thead>
<tr>
<th>Program Component</th>
<th>2019-20</th>
<th>2020-21</th>
<th>2021-22</th>
<th>2022-23</th>
<th>2023-24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Existing Villages</td>
<td>Total Active</td>
<td>9,214</td>
<td>8,002</td>
<td>8,002</td>
<td>4,432</td>
</tr>
</tbody>
</table>

---

2 On average, each village has one government primary school
3 Also referred to as “Primary Partnerships”
Table 2 lists the village types by program component and the evaluation activities that should take place within each.

Table 2: Evaluation activities by village type

<table>
<thead>
<tr>
<th>Program Component</th>
<th>Learning gains impact evaluation</th>
<th>Enrollment verification</th>
<th>Retention verification</th>
<th>Life skills programming verification</th>
<th>School improvement plans verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Existing Villages</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2. AGP</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3. Primary Organic</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>4. Primary Inorganic</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>5. Tech Partnerships</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 lists the number of villages/schools that will be in the impact evaluation sampling frame.

Table 3: Learning program villages and impact evaluation sampling frame

<table>
<thead>
<tr>
<th>Program Component</th>
<th>Unit</th>
<th>2019-20</th>
<th>2020-21</th>
<th>2021-22</th>
<th>2022-23</th>
<th>2023-24</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Primary Organic</td>
<td>Enrollment villages</td>
<td>Total Active</td>
<td>1,000</td>
<td>2,000</td>
<td>4,000</td>
<td>5,000</td>
</tr>
<tr>
<td></td>
<td>New: Year 0</td>
<td>+1000</td>
<td>+1000</td>
<td>+2000</td>
<td>+1000</td>
<td>+1000</td>
</tr>
</tbody>
</table>

4 Although AGP villages are a subset of Existing Villages, Educate Girls counts them separately as part of the 35,000 villages described in the Audacious grant because they represent an exit from the Primary program and the introduction of a new program.

5 While the evaluator should include some Existing Villages in the sample for enrollment, retention, and life skills programming verification, we recommend focusing relatively more resources on evaluation activities in the other four program components, since this will provide more actionable data to inform program updates and course correction.
III. EVALUATION

IDinsight seeks to fund an evaluator to generate evidence on whether Educate Girls successfully delivers the activities described above and on whether the learning component has causal impact on student outcomes. This evidence will be used to track Educate Girls’ progress toward the Audacious targets and inform course correction in the program. This evidence will come from four evaluation activities:

1. An impact evaluation to estimate the causal effect of Educate Girls’ scale-up on learning outcomes in a representative sample of villages that receive the primary learning program
2. Verification data collection to validate Educate Girls’ monitoring data on enrollment and retention in a representative sample of all program villages
3. Verification data collection to validate Educate Girls’ monitoring data on the implementation activities in a representative sample of villages that receive life skills programming
4. Verification data collection to validate the submission of school improvement plans in a representative sample of villages that are included in the school governance and infrastructure improvement programming

Educate Girls will provide lists of program villages, schools, out-of-school girls, and other data in a timely manner that is necessary for the evaluator to conduct these activities. At the same time, the evaluator should seek to minimize the logistical and reporting burden on Educate Girls as much as possible to allow Educate Girls to focus their effort and resources on meeting the Audacious targets.
1. IMPACT EVALUATION OF LEARNING GAINS

Research questions informed by this activity:

1. What is the causal impact of Educate Girls’ program in Primary Organic and Primary Inorganic villages on student learning?
2. How does this impact change over the course of the funding period?
3. How can Educate Girls leverage insights from the impact evaluation to course-correct if necessary, such that Educate Girls is able to deliver on the Audacious Project targets?

The impact evaluation should deliver rigorous causal estimates of the effect of the learning program in Primary Organic and Primary Inorganic program schools. The impact evaluation should collect student data from a representative sample of schools that receive the learning program (as enumerated in Table 3) and control schools, excluding Existing Villages that have already been administering the learning program.

The ideal impact evaluation will provide annual estimates of impact, including one-year effects from schools that are in the first year of implementing the learning program and two-year effects from schools that are in the second year of implementing the learning program. However, recognizing the costliness of large-scale primary data collection, we will also consider evaluation designs that provide results for some but not all years and for some but not all yearly effects. If the evaluator proposes a design that does not deliver annual estimates and multi-year effects each year, the proposal should clearly state how the evaluation will provide adequate and timely evidence for course correction.

One recommended option is for the evaluator to collect outcome data at two points in time:

i. At the end of the 2020-21 school year, or one year after the start of the learning program in Year 1 Primary Organic and Primary Inorganic schools. This data would generate critical evidence on early year impact effects and inform necessary programmatic course corrections.

ii. At the end of the 2022-23 school year, at which point different schools would have received the program for one, two, or three years. This data would generate evidence on the evolving impact of the program over one vs two years, and inform programmatic discussions heading into the final year of the funding period.

IDinsight will work closely with the evaluator to finalize the evaluation design and data collection protocol, drawing on experiences and lessons from the DIB evaluation. Below we suggest specific inputs for the impact evaluation, which are subject to change as information about the program and evidence needs become available. The evaluator is welcome to propose alternative design parameters as long as the evaluation objectives are met within the budget envelope.

A. Evaluation methodology

The evaluator is encouraged to consider an evaluation design that leverages randomized phase-in of the learning program within FC clusters. Each FC cluster will consist of approximately 8 to 10 government primary schools. After a year of preparation activities ("Year 0"), including the D2D survey

---

6 Given the planned roll-out of the program within FC clusters, as described in the following section, it will likely not be possible to measure three-year causal effects.
and Team Balika recruitment and training, Educate Girls will begin implementing the learning program in two schools in an FC cluster starting in Year 1. In Year 2, Educate Girls will roll out the learning program to two additional schools, and in Year 3, Educate Girls will roll out the learning program to two more school, for a total of six schools treated per FC cluster by the end of Year 3.7 Educate Girls will select the six schools to receive the learning program from among all schools in the FC cluster based on school need and certain conditions that are necessary for implementation of the learning program.8 However, Educate Girls has indicated that the evaluator could randomize the order in which these six schools receive the learning program, that is, which two schools receive the program in Year 1, Year 2, and Year 3. The one-year causal impact of the learning program can therefore be estimated by comparing outcomes in Year 1 schools with outcomes in Year 2 & 3 schools at the end of Year 1, and the two-year causal impact can be estimated by comparing outcomes in Year 1 schools with outcomes in Year 3 schools at the end of Year 2.

B. Sampling

It may not be necessary for the evaluator to assess students in all program schools. The evaluator should recommend a sampling strategy that would yield sufficient precision to estimate meaningful treatment effects on learning outcomes for assessment year. As a benchmark, Table A1 in Appendix 2 summarizes standardized effects on test scores for each grade cohort after the first two years of the DIB evaluation.9 Based on this, we recommend designing the evaluation to have sufficient statistical power to detect treatment effects comparable to one-year effect sizes in the first and second years of the DIB evaluation, i.e. ~0.1 SD.

Using the DIB evaluation data, we present several parameters that the evaluator may find useful for calculating the recommended sample size. We present a range of values for each parameter depending on whether one uses Year 1 or Year 2 endline data from the DIB evaluation. The evaluator is welcome to use alternative values provided they are sufficiently justified:

- Autocorrelation in test scores10:
  - One-year autocorrelation range: 0.82 to 0.87
  - Two-year autocorrelation range: 0.70 to 0.81
- Intra-cluster correlation of test scores (within schools):
  - Endline scores: 0.17 to 0.22
  - Endline residuals controlling for baseline scores: 0.11 to 0.19

The evaluator may recommend an evaluation design that samples some schools from all FC clusters or that first samples FC clusters and then samples schools within FC clusters. If the evaluator

---

7 Educate Girls may also consider rolling out the learning program to two more schools in Year 4, for a total of eight treated schools per FC cluster by the end of Year 4.
8 If the evaluator samples a subset of FC clusters for evaluation activities, Educate Girls will specify the six schools that will receive the learning program in sampled clusters in Year 0. For non-sampled FC clusters, Educate Girls may not select schools all six schools in Year 0 to allow for more implementation flexibility.
9 Treatment effects were considerably larger in the 3rd and final year of the DIB. However, since it will not be possible to estimate three-year effects in this impact evaluation, and since we desire sufficient precision to measure small effects than those achieved in the 3rd year of the DIB, we recommend using the 1st and 2nd year DIB effect sizes as benchmarks.
10 Note that the DIB evaluation assessed the same students each year.
recommends the latter design, then we suggest stratifying the sample by Primary Organic and Primary Inorganic villages to ensure sufficient samples from each program sub-component. The evaluator may also use administrative data, such as publicly-accessible data on annual school enrollment and facilities from the District Information System for Education to further improve the efficiency of sampling and/or randomization.

Educate Girls’ learning program targets students in Grades 3, 4, and 5, and so the evaluator should assess students from all of these grades within selected schools. The evaluator is encouraged to consider how the student population will change between baseline and endline and sample students appropriately. For instance, if the learning program leads certain students to enroll or drop out who would not enroll or drop out in the absence of the learning program, sampling students from school registers at endline may lead to biased impact estimates. One option is to sample students from baseline school registers; the evaluator may consider doing the same in this evaluation or may recommend an alternative strategy.

An important subgroup of interest in the impact evaluation is newly-enrolled girls: the evaluator should plan to sample and assess girls who are enrolled over the course of the evaluation from Educate Girls’ list of out-of-school girls. Since Educate Girls will have enumerated all out-of-school girls in the FC cluster before implementing the learning program in Year 1 schools, the evaluator will be able to draw comparable samples of out-of-school girls at baseline from Year 1 and Year 3 school catchment areas.

The evaluator should seek to minimize attrition across the study and avoid differential attrition between treatment and control groups. IDinsight will consider attrition below 25% acceptable, as long as there are no significant differences in the rates of attrition or characteristics of students who attrite across treatment and control groups. However, if either the rate of attrition exceeds 25% or differential attrition occurs, the evaluator will be required to attempt to assess more students until these conditions are no longer violated. The evaluator is encouraged to anticipate attrition through frequent real-time checks during data collection and, if necessary, institute additional testing days at school and/or sample attrited students to be assessed at home.

The evaluator should also propose how spillovers across treatment and control schools will be mitigated, and how crossovers of students between treatment and control schools will be treated in the analysis to avoid bias in impact estimates.

**C. Metrics**

The evaluator should assess students in foundational math and language skills. During evaluation design, the evaluation Working Group (consisting of the evaluator, IDinsight, Educate Girls, and the Audacious funding partners) will determine whether it is necessary to assess Hindi, English, or both. In either case, we encourage the evaluator to consider using the ASER assessment, which is a simple, well-regarded student assessment tool in India with versions freely available online. However, the evaluator should pilot the assessment before data collection and modify it as necessary to capture learning effects in the target population. For instance, it may be necessary to add one or more levels to mitigate ceiling effects, depending on the baseline learning levels of the applicable student population, or make other appropriate modifications as determined by piloting.
D. Data collection

The evaluator should propose annual data collection or an alternate timetable that allows for sufficient evidence to be used to track progress and inform course correction, such as the recommended alternative timeline described at the beginning of this section. The evaluator is encouraged to collect data that would allow for measurement of one-year and two-year effects. Baseline data, if collected, should be collected before the implementation of the learning program in July. While a baseline assessment may not be strictly necessary to estimate causal effects (at least, if the evaluation design describe above is used), it may improve the precision of impact estimates. Endline data should be collected in late February through end of March each year to coincide with the end of the school year while avoiding school exams or other events or holidays in the school calendar.

The evaluator is encouraged to maximize efficiency across data collection activities. For instance, the evaluator might consider collecting baseline and/or endline data at the same time as verification of certain indicators described below. Some or all of the villages in the impact evaluation may provide some or all of the necessary sample for verification data collection in Primary Organic and Primary Inorganic villages.

The evaluator should expect to share primary data with IDinsight and Educate Girls after each round of data collection. The data should exclude any variables that can be used to identify individual children but should include school identifiers. This will enable Educate Girls to use this data to inform course corrections.

2. VERIFICATION OF ENROLLMENT AND RETENTION OF OUT-OF-SCHOOL GIRLS

Research questions informed by this activity:

1. What is the error rate on the enrollment and retention numbers of out-of-school girls reported by Educate Girls?

The evaluator should validate enrollment and retention by collecting data from a representative sample of all program villages and comparing it against Educate Girls’ monitoring data.\footnote{These terms of reference do not include an evaluation of Educate Girls’ causal effect on enrollment or retention for three reasons. First, the cost of conducting a door-to-door census of out-of-school girls in control villages that mirrors the process undertaken by Educate Girls is likely to be prohibitively high. Second, evidence from the DIB evaluation suggests that the counterfactual of enrollment outcomes of out-of-school girls in villages without Educate Girls programming is likely close to zero: The difference in enrollment in treatment and control schools at the year 3 endline was nearly the same as the number of verified girls enrolled by Educate Girls. Third, it would be challenging to identify a robust counterfactual since village selection for enrollment and retention activities is non-random and in fact intentionally selected to maximize program impact.} We include both of these objectives under one activity since we anticipate that it may be most efficient to collect this data from a single representative sample of villages. The evaluator is welcome to propose alternative approaches that use different village samples for each indicator as long as the proposal meets the evaluation objectives.
“Enrollment” is defined as whether a girl who was on Educate Girls’ out-of-school girls lists is added to a school’s enrollment register. In March/April of each year, Educate Girls will supply the evaluator with an updated list of newly-enrolled girls for all sampled villages.12 “Retention” is defined as whether a newly-enrolled girl sits for the final exam at the end of the school year in which she is enrolled. Due to the school summer holiday and Educate Girls’ programming schedule, Educate Girls will not be able to compile the list of retained girls until June/July and will supply the finalized list to the evaluator in September/October. However, it may be efficient for the evaluator to verify enrollment and collect images of exam registers from selected villages at the same time after the summer holiday (e.g. in June/July), and verify retention using exam register images once Educate Girls supplies their list of retained girls.

As with the impact evaluation, the evaluator would ideally verify these indicators each year of the project (2020 through 2024). However, IDinsight will consider designs that provide results for some but not all program years. If the evaluator proposes a design that does not deliver annual estimates, the proposal should clearly state how the evaluation will provide adequate and timely evidence for course correction.

We recommend stratifying village selection by program component (Existing Village program, AGP, Primary Organic, Primary Inorganic, and Technical Partnerships) to ensure sufficient sample from each and to leverage the impact evaluation sample from Primary Organic and Primary Inorganic villages. However, we suggest one exception: the evaluator should consider sampling a smaller fraction of Existing Villages since this data will be mainly be used for accountability purposes, whereas data from program components in expansion regions will provide an important learning value for course correction.

A. Enrollment

The evaluator should sample sufficient villages and girls from Educate Girls’ list of newly enrolled girls to generate precise estimates of the fraction of ineligible girls on the list (the “error rate”). Using this error rate, the Working Group will assess Educate Girls’ performance against predefined thresholds for good, adequate, and poor performance.13 The evaluator should propose a margin of error that is appropriate for the evidence needs of this verification and devise a sampling strategy that will achieve that margin of error.

The Working Group will define eligibility criteria for newly-enrolled girls to be included on Educate Girls’ lists. Possible criteria may include the following, though the final criteria are subject to Working Group discussions: (i) the girl was out-of-school prior to the start of enrollment activities in that village, (ii) she was within the eligible age range (5 to 14 years) at the start of enrollment activities, and (iii) she was successfully enrolled into school. The evaluator should check all of these conditions when verifying eligibility status. It will likely be most efficient for the evaluator to collect previous and

---
12 Educate Girls can provide a draft list of newly enrolled girls in February, though Educate Girls’ quality assurance protocol will likely extend through March. Depending on the timeline and ordering of data collection activities that is most efficient, the evaluator is free to propose to either verify the draft list concurrently with Educate Girls’ quality assurance and reconcile the two lists afterwards, or wait for Educate Girls to complete quality assurance and verify the finalized list afterwards.
13 It is likely that an error rate below 15% will be defined as good, with adequate and poor thresholds being defined at a later date.
current enrollment registers at the same time to check prior and current enrollment status for newly-enrolled girls.

The evaluator should also expect to discuss eligibility discrepancies for individual girls with Educate Girls and, where appropriate, reconcile and update the data. The final error rate will be calculated based on any girls that continue to be listed as ineligible after this reconciliation process. We recommend that the evaluator plan to spend time with Educate Girls’ field staff ahead of data collection rounds to align on data collection strategies for enrollment, retention, and other indicators being verified by the evaluator.

The evaluator should be prepared to verify a relatively larger list of newly enrolled girls at the end of the second year of enrollment activities in a new village (“Year 2”), since historically Educate Girls has enrolled the most girls during this year of programming. After this year, it is likely that the number of newly enrolled girls in that village will fall. However, the evaluator should still be prepared to sample villages after Year 2 of enrollment activities to verify ‘additions’ to the list.

B. Retention

“Retention” is defined as whether a newly-enrolled girl sits for the final exam at the end of the school year in which she is enrolled. In September/October each year, Educate Girls will provide the evaluator with a list of retained girls which the evaluator can verify against school exam registers from the end of the previous year. As noted above, the evaluator may find it most efficient to collect exam registers from selected villages while verifying enrollment, and cross-check those with Educate Girls’ list once it is received. However, the evaluator is welcome to suggest alternative methodologies so long as they meet the objectives of the evaluation. As with the newly enrolled girls list, the evaluator should estimate the error rate of Educate Girls’ list of retained girls list within a reasonable margin of error.

3. VERIFICATION OF LIFE SKILLS EDUCATION

Research questions informed by this activity:

1. Do life skills programming activities occur as reported by Educate Girls?

The evaluator should validate the delivery of life skills programming by collecting data from a representative sample of program villages that receive life skills education (Primary Organic, Primary Inorganic, Existing Villages, and AGP) and comparing it against Educate Girls’ monitoring data. To minimize data collection costs, the evaluator is encouraged to leverage the sample of villages where enrollment and retention verification will take place, excluding those villages where life skills programming will not be delivered (Technical Partnership villages). As with enrollment and verification activities, we encourage the evaluator to sample a larger fraction of villages in expansion areas (Primary Organic, Primary Inorganic, and AGP villages), since this data will be more valuable for course-correction.

14 These terms of reference do not include an impact evaluation of the life skills programming due the difficulty of identifying a rigorous counterfactual for non-randomly selected AGP villages and due to budgetary constraints to collecting this data in control primary villages. It is likely that Educate Girls will commission an internal impact evaluation of life skills programming at some point during the Audacious funding timeline.
The evaluator should propose activities that will verify implementation of life skills education as cost-effectively as possible. This may include interviewing students involved in the Bal Sabhas or Kishori Samuhs, in-person spot checks of program implementation, and/or document verification. The evaluator should propose metrics and a sampling strategy that will yield reasonably precise estimates of the fidelity of implementation.

4. VERIFICATION OF SCHOOL IMPROVEMENT PLAN SUBMISSIONS

Research questions informed by this activity:

1. Are school improvement plans submitted as reported by Educate Girls?

The evaluator should validate the submission of school improvement plans by collecting data from a representative sample of program villages that receive school governance and infrastructure support (Primary Organic, Primary Inorganic, and Existing Villages) and comparing it against Educate Girls’ monitoring data. To minimize data collection costs, the evaluator is encouraged to leverage the sample of villages where enrollment and retention verification will take place, excluding those villages where school governance and infrastructure support will not be provided (Existing Villages, AGP, and Technical Partnership villages).

Each year, Educate Girls will submit a list of villages that have prepared school improvement plans, and the evaluator should verify the presence of these plans in a representative sample. The evaluator may consider cost-effective alternatives for data collection, such as reviewing digital images of school improvement plans and interviewing headmasters over the phone. More details on the content and format of these improvement plans will be provided by Educate Girls prior to data collection.

5. ETHICAL CLEARANCE

The evaluator is expected to secure ethical clearance from an accredited institutional review board (IRB) for all data collection activities, and to abide by the rules specified in IRB clearance regarding informed consent, data security, and other ethical research practices. The evaluator should seek IDinsight’s approval before finalizing selection of the proposed IRB.

6. DELIVERABLES

At a minimum, the evaluator should report results from the impact evaluation and verification activities on an annual basis. If the evaluator proposes a design that does not include data collection every year, then the evaluator should submit an updated workplan during ‘off’ years. The evaluator should expect to stay in frequent communication with IDinsight to report progress, inform about problems encountered, and obtain approval for deviations from the evaluation plan.

---

15 The terms of reference do not include an impact evaluation of school improvement plan submission since these plans would not be prepared in the absence of Educate Girls staff and Team Balika volunteers.
Reporting formats should include concise written reports and accompanying sliddecks. The evaluator should expect to iterate on reports with IDinsight before submitting revised drafts and presenting sliddecks to the Working Group. The evaluator will also be required to share cleaned datasets and statistical analysis code with IDinsight.

7. EVALUATION BUDGET

The budget set aside for all evaluation activities, including design, data collection, analysis, and reporting for the impact evaluation and verification activities is USD 400,000. Any proposals that exceed this budget envelope will not be accepted.

The evaluator may recommend a payment schedule that corresponds to the proposed timeline of evaluation activities and deliverables. IDinsight will not pay more than 20% of the total value of the contract before the first deliverable is accepted by IDinsight, and will reserve at least 20% of the total value of the contract for the acceptance of the final deliverable.

The evaluator is encouraged to identify creative ways to deliver on the evaluation objectives as cost-effectively as possible. These may include but are not limited to:

- Sampling a representative sample of districts, FC clusters, villages, schools, and students, rather than collecting data from all program districts, clusters, villages, schools, and students.
- Combining data collection rounds for multiple evaluation activities, such as collecting the previous year’s enrollment and retention verification data at the same time, or administering endline learning assessments at the same time as verifying the current year’s enrollment data, or administering baseline learning assessments at the same time as verifying the previous year’s retention data.
- Leveraging Educate Girls data to facilitate verification (without fully substituting for evaluator’s primary data).
- Leveraging administrative data to improve precision for the impact evaluation or supplement verification activities.
- Collecting data remotely (such as through pictures supplied by school officials or phone interviews), if appropriate.
- Collecting data in some but not all years, such as at the ends of the 2020-21 and 2022-23 school years, such that the evidence from the evaluation can be used for course-correction and for measuring one-year and two-year effects of the program (though not necessarily from the same schools).

IV. EVALUATOR SELECTION PROCESS

1. ANTICIPATED EVALUATOR SELECTION TIMELINE

The following timeline is subject to change. IDinsight will inform prospective bidders of any changes.

- October 14, 2019: Release of TOR
• **October 31, 2019**: Prospective bidders submit questions about the proposal or evaluation to IDinsight
• November 4, 2019: IDinsight shares responses to questions with prospective bidders
• **November 18, 2019**: Proposals due from bidders at 23:59 Pacific Time
• December 2 – December 6, 2019: Phone interviews with shortlisted bidders
• December 10, 2019: The successful bidder is selected and announced
• December 10, 2019 – February 29, 2020: The evaluator finalizes the evaluation design in close collaboration with IDinsight; prepares field protocols and data collection instruments; registers the evaluation on a public registry and submits a pre-analysis plan; prepares field teams for data collection; samples eligible villages and schools for the impact evaluation and verification activities; and randomizes eligible schools for the impact evaluation.
• March 2020: The evaluator conducts the baseline learning assessment and the first round of verification

2. **PROPOSAL REQUIREMENTS**

Proposals should not exceed 10 pages (excluding appendices), should be submitted as a .docx file, and should include the following items:

- Recommended design for each of the four evaluation activities described above, including preliminary sampling strategies and sample size calculations.
- Plan for data collection, including metrics, frequency of data collection (referencing the evidence needs of the Working Group), field team structure, and contingencies for non-compliance, attrition, and other potential evaluation issues.
- Proposed framework or models for data analysis.
- Timeline and workplan for activities.
- Budget in USD broken down by each evaluation activity and explaining cost assumptions (including daily rates).
- Team structure (team leader and key personnel) with one paragraph describing the relevant qualifications and the proposed level of effort, as a percentage of full-time equivalent, for each team member. Full CVs should be added to the appendix.
- Professional qualifications, including previous relevant experience, of the bidder/organization. Attach up to 3 reports or publications (ideally topically and/or geographically relevant) that demonstrate the bidder’s experience in applied empirical methods as separate documents.

The submission must be clear, concise and complete. IDinsight reserves the right to mark a submission down or exclude it from the process if its submission contains any ambiguities or lacks clarity. Applicants should submit only such information as is necessary to respond effectively to this request for proposals. Unless specifically requested, extraneous presentation materials are neither necessary nor desired. Submissions will be evaluated on the basis of information submitted by the deadline.

Please send questions and proposals to jeffery.mcmanus@idinsight.org, clearly stating “[Questions or Proposal] for Evaluation of Educate Girls Program – [name of bidder/agency]” as the subject line.
Questions must be submitted by **11:59pm (Pacific Time) on October 31, 2019**, and full proposals must be submitted by **11:59pm (Pacific Time) on November 18, 2019**.

### 3. CRITERIA FOR EVALUATOR SELECTION

Proposals will be assessed on the following criteria. Proposals that exceed the budget allocated for this evaluation will not be accepted:

- Quality of the technical proposal, including its ability to meet the evaluation objectives within the constraints of the program and context described above.
- Experience of the organization and key personnel in designing rigorous evaluations and collecting data relevant to the evaluation activities.
- Value for money in terms of the scope and frequency of data collection and evidence generated within the budget envelope.

Shortlisted candidates will be given an opportunity to present their proposal during a one-hour phone interview the week of December 2, 2019. The successful bidder will be informed on or around December 10, 2019, and evaluation design is expected to begin immediately.

### V. APPENDIX

#### 1. OVERVIEW OF ORGANIZATIONS

**Educate Girls** is a non-profit that focuses on mobilising communities for girls’ education in India’s rural and educationally backward areas. Strongly aligned with the ‘Right to Education Act’ or the ‘Samagra Siksha’, Educate Girls is committed to the Government’s vision to improve access to primary education for children, especially young girls. Educate Girls currently operates successfully in over 13,000 villages in Rajasthan and Madhya Pradesh. By leveraging the Government’s existing investment in schools and by engaging with a huge base of community volunteers, Educate Girls helps to identify, enroll and retain out-of-school girls and to improve foundational skills in literacy and numeracy for all children (both girls and boys). This helps deliver measurable results to a large number of beneficiaries and avoids duplication or parallel delivery of services.

**IDinsight** uses data and evidence to help leaders combat poverty worldwide. Our collaborations deploy a large analytical toolkit to help clients design better policies, rigorously test what works, and use evidence to implement effectively at scale. We place special emphasis on using the right tool for the right question, and tailor our rigorous methods to the real-world constraints of decision-makers. IDinsight works with governments, foundations, NGOs, multilaterals and businesses across Africa and Asia. We work in all major sectors including health, education, agriculture, governance, digital ID, financial access, and sanitation. We have offices in Bengaluru, Dakar, Johannesburg, Lusaka, Manila, Nairobi, New Delhi, San Francisco, and Washington, DC.

Educate Girls and IDinsight are collaborating on a learning partnership to generate data and evidence that will inform Educate Girls’ programmatic decisions during the five-year scale-up. This partnership will include (i) generating and employing predictive algorithms to target the highest-need areas; (ii)
building robust monitoring systems for each of Educate Girls’ sub-programs; and (iii) conducting evaluations of pilot initiatives, among other activities. Given this ongoing collaboration, IDinsight has opted to not conduct the external evaluation activities described in these terms of reference. However, IDinsight will work closely with the evaluator to ensure the appropriateness and robustness of all evaluation designs and analyses, drawing on IDinsight’s experience conducting the DIB evaluation and leveraging ongoing relationships with key stakeholders at Educate Girls.

The Children’s Investment Fund Foundation (CIFF) is an independent philanthropic organisation, headquartered in London with offices in Nairobi and New Delhi. CIFF works to transform the lives of poor and vulnerable children in developing countries. CIFF’s areas of work include children and mothers’ health and nutrition, children’s education and welfare and smart ways to slowdown climate change. CIFF-funded programmes place significant emphasis on quality data and evidence. Before making an investment and during implementation, CIFF works with partners to measure and evaluate progress to achieve large scale and sustainable impact.

The Audacious Project is a collaborative approach to funding big ideas with the potential to create change at thrilling scale. Housed at TED, which has a long track record of surfacing ideas worth spreading, and supported by The Bridgespan Group, the project is a unique partnership between some of the most respected organizations in philanthropy and the public. The Audacious Project attempts to address a major frustration faced by the world’s change-makers. Without access to venture capital or stock markets, social entrepreneurs have to pitch donors one-by-one, often a deeply inefficient process for all involved. The Audacious Project aims to: (1) encourage the world’s greatest change-agents to dream bigger than ever before (2) shape their best ideas into viable multi-year plans and (3) present those ideas in a compelling way to potential supporters. The project’s goal is to make philanthropy more collaborative, more inspiring — and more effective.

2. EXISTING EVIDENCE: DEVELOPMENT IMPACT BOND EVALUATION

Educate Girls participated in the world’s first development bond (or “DIB”), the Educate Girls DIB, from 2015 to 2018. The DIB set three-year impact targets for enrollment of out-of-school girls and learning gains of boys and girls in grades 3 to 5 in Bhilwara District, Rajasthan. Under the contract terms, the UBS Optimus Foundation disbursed payments to Educate Girls in the first year of programming and was repaid by the Children’s Investment Fund Foundation proportional to Educate Girls’ success against targets at the end of the third year. A subcontract between Educate Girls and the UBS Optimus Foundation further stipulated that the UBS Optimus Foundation would pass on an incentive payment to Educate Girls for any return on their investment. More details on how the targets were set, the financial structure of the DIB, and the roles of different stakeholders are provided on the Educate Girls DIB website and in the DIB design memo.

As part of this DIB, IDinsight conducted a three-year randomized controlled trial of Educate Girls program across 332 schools and 282 villages in Bhilwara District. Half of villages in the study sample were randomly assigned to receive Educate Girls’ program while the other half formed the comparison group. IDinsight tracked a cohort of 12,000 students in grades 3-5 over the course of the 3-year evaluation and conducted annual assessments of students’ Hindi literacy, numeracy, and English skills. IDinsight also verified enrollment of out-of-school girls in treatment villages, though due to data collection costs, the DIB Working Group decided against tracking out-of-school and newly-enrolled girls in control villages. 80% of DIB payments were based on Educate Girls’ causal impact on learning.
outcomes, as measured through the RCT, and the remaining 20% were based on the percent of out-of-school girls in treatment villages whose enrollment was verified by IDinsight.

At the end of the three-year DIB, Educate Girls had surpassed the enrollment target by 16% and the learning target by 60%, triggering the maximum return on UBS Optimus Foundation’s investment. While Educate Girls had a positive but relatively modest effect on learning outcomes in the first two years of the program, the third year of the program saw a large increase in treatment effects, with learning gains comparable to the most effective remedial education interventions rigorously evaluated in India. Students in Educate Girls schools during the last year of the program gained 0.55 standard deviations relative to control students, representing an additional 1.2 years of schooling in Hindi, Math, and English. IDinsight’s final technical report provides more details on the design and results of the DIB evaluation.

Table A1 specifies treatment effects (in terms of standard deviations of the control group) for each subject tested in the first two years of the DIB evaluation. We omit the three-year DIB effects from this table since the one-year and two-year effects are more relevant for the evaluation described in this ToR; due to Educate Girls’ planned phase-in within FC clusters, it will likely not be possible to measure three-year effects. Moreover, the three-year DIB effects were considerably larger than the one-year and two-year effects, and so the desired minimum detectable effect for this evaluation is closer to the one-year and two-year effect sizes. For three-year results, please refer to IDinsight’s final technical report.

Table A1: Standardized treatment effects by cohort and subject for the first two years of the DIB evaluation

<table>
<thead>
<tr>
<th>Cohort (Grade in Year 1)</th>
<th>Math</th>
<th>Hindi</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year 1</td>
<td>Year 2</td>
<td>Year 1</td>
</tr>
<tr>
<td>1</td>
<td>Not tested</td>
<td>Not tested</td>
<td>Not tested</td>
</tr>
<tr>
<td>2</td>
<td>Not tested</td>
<td>0.073*</td>
<td>Not tested</td>
</tr>
<tr>
<td></td>
<td>[0.044]</td>
<td>[0.043]</td>
<td>[0.051]</td>
</tr>
<tr>
<td>3</td>
<td>0.046</td>
<td>0.150***</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>[0.035]</td>
<td>[0.038]</td>
<td>[0.037]</td>
</tr>
<tr>
<td>4</td>
<td>0.084**</td>
<td>0.192***</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>[0.039]</td>
<td>[0.048]</td>
<td>[0.032]</td>
</tr>
<tr>
<td>5</td>
<td>0.168***</td>
<td>Not tested</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>[0.034]</td>
<td>[0.027]</td>
<td>[0.033]</td>
</tr>
<tr>
<td>Pooled</td>
<td>0.108***</td>
<td>0.150***</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>[0.023]</td>
<td>[0.028]</td>
<td>[0.018]</td>
</tr>
</tbody>
</table>

Each student cohort is defined by the grade that cohort was in during Year 1. Only Grades 3, 4, and 5 were tested each year, and so some cohorts were omitted from testing in certain years. Each cell represents the coefficient on treatment in a regression specification for that cohort at the end of each year. Standard errors clustered at the village-level are in brackets below coefficients. * p < 0.1, ** p < 0.05, *** p < 0.01