

# **Educate Girls Development Impact Bond** Final Evaluation Report

10 June 2018

In Partnership with: Children's Investment Fund Foundation, Educate Girls, Instiglio, and UBS Optimus Foundation



# **Table of Contents**

5
5
7
11
11
12
12
13
14

# About the Author

IDinsight helps clients generate and use rigorous evidence to improve social impact. Depending on client needs, we help diagnose social sector challenges, design and test potential solutions, and operationalize those solutions found to be most impactful. We believe that client-centered, rigorous, and responsive evaluation is essential to help managers maximize program impact.

# About the Educate Girls Development Impact Bond

The Educate Girls Development Impact Bond (EG DIB) is a joint project between the Children's Investment Fund Foundation (CIFF), Educate Girls (EG), the UBS Optimus Foundation, Instiglio, and IDinsight (collectively, the "Working Group") to provide and improve education for girls in rural India. UBS Optimus, acting as the investor, financed EG's project implementation, while CIFF will pay for educational outcomes as evaluated by IDinsight. Instiglio is managing the project.



# A Note on Grade and Student Cohort Labels

Over the course of the three-year evaluation, IDinsight tracked five different grades of students as they progressed through school. At Baseline, we assessed students in grades 1 through 5. In each subsequent Endline, we assessed students who were then in grades 3, 4, and 5 (the target grades for Educate Girls' programming). Since a student's grade changes year to year, student cohort labels can be ambiguous; for instance, "Grade 3" could refer to three different cohorts of students in the evaluation (students who were  $3^{rd}$  graders in Year 1, Year 2 or Year 3 of the evaluation). To remove this ambiguity, in this report **we refer to student cohorts according to their grade in Year 1**, unless explicitly noted otherwise. We attach the "Y1" suffix to grade labels to remind the reader of this convention. For instance, " $2_{Y1}$ " refers to students in grade 2 during the first year of the evaluation, who had progressed to grade 3 in Year 3.

Table 1 shows how each cohort progressed through school during the evaluation and how many years students in the treatment group were potentially exposed to EG programming. Gray cells indicate when the cohort was assessed by IDinsight.

Student cohort label	Grade le	vel at each	Years of exposure to EG program		
	Baseline	Y1 Endline	Y2 Endline	Y3 Endline	
Grade 1 <sub>Y1</sub>	1	1	2	3	1
Grade 2 <sub>Y1</sub>	2	2	3	4	2
Grade 3 <sub>Y1</sub>	3	3	4	5	3
Grade 4 <sub>Y1</sub>	4	4	5	6	2
Grade 5 <sub>Y1</sub>	5	5	6	7	1

**Table 1: Student Cohorts During the Evaluation** 



# **1. Executive Summary**

In this report, we present the results of IDinsight's three-year impact evaluation of Educate Girls' program in Bhilwara District in Rajasthan, India. The two outcomes described in this report – learning gains of students enrolled in grades 3-5 and enrollment of out-of-school girls – will determine the payments in the Educate Girls Development Impact Bond.<sup>1</sup> Educate Girls surpassed the DIB targets for both learning gains and enrollment.

#### **Outcome 1: Learning gains**

By the end of Year 3, students in treatment villages gained an additional 8,940 ASER learning levels relative to students in control villages, representing 160% of the final target.<sup>2</sup>

*Methodology*: IDinsight conducted a three-year clustered randomized controlled trial in which we compared students in schools where EG operated with students in control schools. We assessed students on basic literacy and math competencies using the Annual Status of Education Report (ASER) testing tool; a student's score on ASER determined her "learning level," which is scored out of 16 points and forms the basis of the learning metric.

*Results*: On average, students in EG schools gained an additional 1.08 ASER levels compared to students in control schools (p < 0.01). Differences in aggregate learning gains between treatment and control schools were much greater in Year 3 (+6,045 learning levels) than in Year 2 (+1,434 levels) or in Year 1 (+1,461 levels).<sup>3</sup>

#### *Outcome 2: Enrollment*

# By the end of Year 3, Educate Girls enrolled 92% of all 837 eligible out-of-school girls in treatment villages,<sup>4</sup> representing 116% of the final target for enrollments.

*Methodology:* IDinsight used a simple pre-post comparison to verify enrollments of out-ofschool girls in treatment villages. Due to the cost of conducting a census of all households, the Working Group decided against estimating enrollments in control villages. Thus, unlike the learning estimates, the enrollment estimates do not reflect a causal effect of EG's program.

*Results:* Educate Girls enrolled 155 girls in Year 3, or 19% of all eligible out-of-school girls identified since the start of the evaluation. Including the 613 enrollments from Years 1 and 2, Educate Girls enrolled a total of 768 out of 837 eligible out-of-school girls.

Outcome	Methodology	Target	Final Result	Performance as Percent of Target
Aggregate learning gains for all students in grades 3-5	Clustered (village-level) randomized controlled trial	+5,592 ASER learning levels above control group gains	+8,940 ASER learning levels above control group gains	160%
Enrollment of out-of-school girls	Pre-post comparison	79% of all eligible out-of-school girls	92% of all eligible out-of-school girls enrolled	116%

<sup>&</sup>lt;sup>1</sup> Approximately 80% of the outcome payments are based on changes in learning levels. Approximately 20% are based on changes in enrollment of out-of-school girls. See the <u>Evaluation Design Memo</u> for a full description of how payments will be calculated.

<sup>&</sup>lt;sup>2</sup> This target was revised down from 6,664 to 5,592 by the Working Group in Year 1. EG would have surpassed the original target by 34%.

<sup>&</sup>lt;sup>3</sup> This is a slight change from the result reported in the Year 2 Endline report (1,314 learning for Year 2, 1,498 for Year 1), reflecting updates to the data made in Year 3 as per Appendix 14.

<sup>&</sup>lt;sup>4</sup> At the beginning of the 2017-2018 academic year, Educate Girls identified 90 additional out-of-school girls along with 88 girls already on the list who had left the area of program coverage. This resulted in a final population of 837 out-of-school girls eligible for enrollment.



# 2. Outcome I: Learning Gains

# Methodology

IDinsight conducted a three-year randomized controlled trial, clustered at the village level, to estimate learning gains attributable to EG's program.<sup>5</sup>

#### Sampling and Randomization

The evaluation was conducted in 332 schools across 282 villages in rural Rajasthan, which were selected according to the process outlined in Figure 1, below.



**Figure 1: Sampling and Randomization Protocol** 

Note: \* Village and school eligibility criteria are based on data in the 2014-15 DISE database unless otherwise indicated. \*\* Students are considered assessed if at least one Endline score is available.

Our study population consists of all students who were enrolled in treatment or control schools at Baseline as well as students who enrolled during the evaluation.<sup>6</sup> In the results

<sup>&</sup>lt;sup>5</sup> See the <u>Evaluation Design Memo</u> for a full treatment of the methodology.

<sup>&</sup>lt;sup>6</sup> For students in grades 1 and 2 at Baseline, we attempted to assess all students in the population. For students in grades 3, 4, and 5 at Baseline, due to budgetary constraints we assessed a random sample of 69% of students in the population, stratified by gender and grade. During analysis we apply appropriate sample weights to these students' outcomes to recover population-level learning gains. For example, if 60% of eligible students in a school-grade-gender cohort were sampled, then their learning gains are multiplied by 1/60% = 1.66 in the final analysis. If 100% of eligible students in a school-grade-gender cohort were sampled (as with all grade 1 and 2 students), then their learning gains were multiplied by 1/100% = 1 in the final analysis.



section below, we present average and aggregate results for the full sample of students unless otherwise indicated. In the appendix, to provide points of comparison with previous reports, we also present results separately for students present at Baseline (also called "Type I-III" in the <u>Design Memo</u>) and students absent at Baseline ("Type IV-V").<sup>7</sup> If students were absent from school on the day of the assessment then we assessed them at home.<sup>8</sup>

We separately report learning gains of newly enrolled girls from EG's out-of-school girl lists, which are included in aggregate learning gains calculations and DIB payments. Since we did not collect comparable data in control villages, we exclude these girls from the average treatment effect results.

The third and final Endline was conducted between February 2 and February 28, 2018 and is described in Appendix 3. Please refer to the Year 1 and Year 2 reports for further details on data collection in those years.

#### **Student Assessments**

Learning gains were measured using the Annual Status of Education Report (ASER) assessment tool (see Table 2 below and Appendix 18). The ASER assessment consists of three sections: Hindi, Math, and English. Each section consists of 5 levels (and a possible score of 1 to 5 points). IDinsight added one additional level to the Hindi section ("Story Plus") to reduce "ceiling effects," in which the highest score on a section underestimates a student's true ability. The highest possible total score on this assessment is thus 16 points (5 + 5 + 6); the lowest possible score is 3 points (1 + 1 + 1).

Level	Hindi	Math	English
1	Beginner	Beginner	Beginner
2	Letters	Numbers 1-10	Capital letters
3	Words	Numbers 11-99	Lowercase letters
4	Paragraph	Subtraction	Words
5	Story 1	Division	Sentences
6	Story Plus	_	_

#### Table 2: Learning Levels as Measured by ASER

#### **Calculating Learning Gains**

The change in learning levels for each student is calculated by subtracting his or her total score at Baseline from his or her total score at Endline,<sup>9</sup> with the following caveats:

- Baseline scores for students in treatment and control schools who were not present at Baseline are imputed to be the lowest score possible (a score of 3) and any additional learning levels achieved by those students at Endline are assumed to be gains.
- Students with no Endline score from any round are not included in the analysis (466 students).

<sup>&</sup>lt;sup>7</sup> While secondary to the full sample results, we believe that distinguishing between students present at Baseline versus absent at Baseline is a useful robustness check. Students who were present at Baseline form a consistent sample throughout the three-year evaluation and are thus comparable between treatment and control schools. On the other hand, students who were absent at Baseline are composed of both students who were absent but enrolled at Baseline and students who enrolled in schools later. Since EG's programming includes enrollment activities, students who were absent at Baseline are not directly comparable between treatment and control schools, limiting our ability to make causal claims about their learning gains.

<sup>&</sup>lt;sup>8</sup> Due to cost and logistical constraints we did not assess students in grade 5 at Baseline who were absent on the day of the assessment and had graduated out of the program after Year 1. Per the Working Group's decision in Year 2, the learning gains of these students were imputed based on the learning gains of students in grade 5 who were present on the day of the Baseline assessment.

<sup>&</sup>lt;sup>9</sup> This is a difference-in-differences estimator. For more information, see the Evaluation Design Memo.



- For students who were assessed during multiple Endlines (for example, students who were in grade 3 during Endline Year 1, grade 4 during Endline Year 2, and grade 5 during Endline Year 5), only the final Endline score is counted.<sup>10</sup>
- We apply sampling weights to each group of students according to the proportion of students selected for assessment from this group.

## **Findings**

We present both *average* treatment effects and *aggregate* treatment effects.<sup>11</sup> Average treatment effects are the difference in average learning gains between treatment and control students,<sup>12</sup> and are particularly useful for understanding the magnitude of the program's impact and comparing it to other interventions. Aggregate treatment effects are calculated by adding up learning gains of all students in treatment schools and subtracting learning gains of all students in control schools, and therefore account for differences in the number of students in treatment and control schools due to EG's enrollment activities and other factors.<sup>13</sup> The final Development Impact Bond payments are based on aggregate treatment effects.

#### Learning Gains against the DIB Target

Students in EG schools gained on average an additional 1.08 ASER learning levels compared to students in control schools (p < 0.01).<sup>14</sup> Learning gains for students in EG schools are 28% or 0.31 standard deviations larger than gains for students in control schools, comparing favorably with primary school programs aimed at improving test scores in rural India.<sup>15</sup>

With these large learning gains, EG exceeded the three-year DIB aggregate treatment effect target. By the end of the three-year program, students in treatment villages had gained an additional 8,940 learning levels relative to students in control villages, representing 160% of the final target of 5,592. Figure 2 shows year-to-year growth in the difference in aggregate learning gains between treatment and control students, with more than two-thirds of the difference occurring in year 3.

<sup>&</sup>lt;sup>10</sup> 33 students who should have graduated out of the program were retained. We assessed these students during their additional retention year and use their final score to calculate learning gains.

<sup>&</sup>lt;sup>11</sup> We present average and aggregate results for the full sample of students unless otherwise indicated. In the appendix, to provide points of comparison with previous reports, we also present results separately for students present at Baseline (also called "Type I-III" in the Design Memo) and students absent at Baseline ("Type IV-V"). We separately report learning gains of newly enrolled girls from EG's out-of-school girl lists, which are included in aggregate learning gains calculations and DIB payments. Since we did not collect comparable data in control villages, we exclude these girls from the average treatment effect results.

<sup>&</sup>lt;sup>12</sup> Technically, we control for Baseline learning levels in a linear regression rather than subtracting Baseline learning levels from Endline learning levels.

<sup>&</sup>lt;sup>13</sup> By using aggregate treatment effects as the DIB payment metric, EG was incentivized to enroll out-of-school students even if their learning levels were very low and would bring down the school average.

<sup>&</sup>lt;sup>14</sup> The difference in learning gains is statistically significant at the 1% level. This means that the probability of observing this difference due to random chance, if the treatment effect is zero, is less than 1%. Since this probability is very low, we reject the null hypothesis that the gains in learning levels were equal in program and control villages. Due to randomization we can reasonably expect that, on average, the only difference between students in treatment villages and students in control villages is that the former have been exposed to Educate Girls' program. Balance checks presented in the Baseline report show that there are no statistically significant differences between the control and treatment groups across any of the variables collected.

<sup>&</sup>lt;sup>15</sup> According to an evidence review of education evaluations in developing countries conducted by the Abdul Latif Jameel Poverty Action Lab, an increase in test scores of less than 0.1 SD is typically considered to be a small effect, while an increase of more than 0.3 SD is considered a large effect, and an increase of more than 0.5 SD a very large effect. Among the programs included in the evidence review, the Balsakhi Program, a remedial tutoring education intervention implemented in schools in Vadodara and Mumbai, may be the most similar to EG's program. In that evaluation, the Balsakhi program increased average test scores by 0.28 standard deviation (Banerjee et al. 2007). The same evaluation found no discernible impact of reducing class sizes on test scores. Other evaluations of primary school programs in rural India have found effects on math and language test scores ranging from 0.16 to 0.47 standard deviations (e.g. Duflo, Hanna, and Ryan 2012; Muralidharan and Sundaraman 2012; Banerjee et al. 2007).



These differences resulted from a combination of increased learning and increased enrollment in treatment schools, though relatively more from learning. By the end of Year 3, our study population included 7,318 students in treatment schools and 6,786 students in control schools, reflecting a modest increase in enrollment due to EG's program. The majority of this difference can be explained by the 421<sup>16</sup> out-of-school girls EG reported enrolling in grades 3 to 5 in treatment schools during the study. Excluding learning gains among these newly-enrolled girls, students in treatment schools gained 7,719 more learning levels than their peers in control schools, representing 86% of the difference in aggregate learning gains between treatment and control schools.

Appendices 10 and 11 provide additional detail about how aggregate learning gains break down across grade and student type.



#### Figure 2: Aggregate Learning Gains (Treatment-Control) by Year

#### Learning Gains by Cohort

Treatment effects vary across grades and years. Figure 3 shows average learning gains for treatment and control students by grade at Baseline.<sup>17</sup> Grade  $1_{Y1}$  refers to students who were in grade 1 at Baseline, Grade  $2_{Y1}$  to students who were in grade 2 at Baseline, and so forth. Each year, EG's program targeted students in grades 3-5. Hence, students in Grade  $1_{Y1}$  entered the program for the first time in Year 3, and Grade  $5_{Y1}$  students exited the program after the first year. Students in Grade  $3_{Y1}$  were the only cohort to receive the program for all three years.

Figure 3 provides two major insights. First, program impact increases with years of program exposure. Students in Grade  $3_{Y1}$ , who were exposed to EG's programming for all three years, had the largest learning gains of any cohort. Second, EG's intervention in Year 3 was far more

<sup>&</sup>lt;sup>16</sup> While girls enrolled from the list of eligible out-of-school girls were counted towards the enrollment target regardless of their grade, their learning gains were only assessed if they were in grades 3-5 at the time of one of the Endline surveys.

<sup>&</sup>lt;sup>17</sup> Figure 3 omits students who were absent at Baseline since these students were only assessed during the Year 2 and Year 3 Endlines. In Appendix 6, 7, and 11 we present final results for all student types.



effective than in previous years.<sup>18</sup> Students who participated in the program in Year 3 benefitted 2-3 times more than their peers who had aged out of the program prior to Year 3. Treatment students in Grades 2<sub>Y1</sub> and 3<sub>Y1</sub> grew an astonishing 79% more during the final year of the program than their peers in control schools.



#### Figure 3: Average Learning Levels by Cohort

Note: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Lines omit students absent at baseline (since they do not have a baseline score), though average treatment effects (ATEs) include all students. ATEs denote the difference in average learning gains between students in program schools and students in control schools.

Figure 4 shows the effect of EG's program on learning gains by project year for each grade targeted by EG (grades 3-5). Hence, each bar denotes the additional learning gains achieved in program schools within that year compared with gains among comparable students in control schools. For instance, the first bar shows the difference in average learning gains (+0.18) for students in Grade 3 during the first year of the program (2015-16, corresponding to cohort  $3_{Y1}$ ), and the second bar shows the difference in average learning gains (+0.09) for students in Grade 3 during the second year of the program (2016-17, corresponding to cohort  $2_{Y1}$ ).

Across all grades, the one-year effects of the program in Year 3 far exceed the effects in previous years. The difference is greatest for students in Grade 3: whereas the program did not have a statistically significant effect on learning gains for Grade 3 students in previous years, in the final year of the program Grade 3 students made gains comparable to older peers.

<sup>&</sup>lt;sup>18</sup> The structure of the Development Impact Bond gave EG the flexibility to revise its teaching intervention throughout the three-year project.





Figure 4: One-Year Average Treatment Effects by Grade and Year

Note: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Range bars denote 95% confidence intervals. Since we did not assess students at the beginning of grade 3 in Year 2 and 3, we calculate grade 3 treatment effects using baseline scores for those cohorts. The one-year comparison therefore assumes that any treatment effects for these cohorts occurred during Grade 3 only. The yearly average treatment effects for each cohort do not sum exactly to the overall average treatment effect for that cohort since the yearly average treatment effects do not account for students who have dropped out or have been retained.

#### Learning gains by subject, gender, and geography

Figure 5 shows average learning gains for all students by subject and treatment status. Program impacts were concentrated in Math and English, where the treatment effects were approximately 3 times larger than in Hindi.<sup>19</sup> Appendix 12 further shows that students with low baseline scores, especially in Math and English, benefitted the most from EG's program.

As in previous years, average treatment effects were larger for students in Bijoliya block than for students in Mandalgarh and Jahajphur. Girls benefitted slightly more than boys (+1.13 vs. +1.04).

<sup>&</sup>lt;sup>19</sup> Appendix 7 shows treatment effects separately for students present at Baseline and absent at Baseline.





Figure 5: Average Learning Gains by Subject and Treatment Status

Note: \* p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01. Average treatment effects (ATEs) denote the mean difference in learning gains between students in program schools and students in control schools. Range bars denote 95% confidence intervals. The figure includes data from all Endlines. For a subject-wise analysis of average treatment for Year 3, see Appendix 8.

# 3. Outcome II: Enrollment of Out-of-School Girls

## Methodology

Educate Girls compiled and maintained a census of out-of-school girls in treatment villages, which IDinsight validated each year according to the process shown in Figure 6. Due to budgetary constraints, the DIB Working Group decided not to conduct a parallel census of out-of-school girls in control villages. As a result, we cannot rule out the possibility that other factors besides the Educate Girls program influenced enrollment in treatment villages.

#### **Figure 6: Enrollment Verification Process**



To validate enrollment each year, IDinsight surveyors visited each school in which a girl was reported enrolled and presented the headmaster with a form that included the girl's name, caste, age, and father's name. Headmasters were requested to verify this information by signing the IDinsight form as well as by showing surveyors the register.



## **Findings**

Figure 7 shows the results of this validation exercise. Including the enrollments from Year 1 and Year 2, EG enrolled 768 out-of-school girls, representing 92% of the 837 eligible<sup>20</sup> out-of-school girls.<sup>21</sup> EG exceeded the enrollment target of 79% by 13 percentage points, or 16%.



Figure 7: Enrollments of Out-of-School Girls by Year

Note: Percentages refer to the percent of enrolled girls relative to the Year 3 target of 837 eligible girls. The list of eligible out-of-school girls was updated each year to include newly-eligible girls and exclude newly-ineligible girls.

# 4. Conclusion

Educate Girls exceeded the 3-year DIB targets in both learning and enrollment. Students in program villages gained an additional 8,940 ASER learning levels relative to comparable students in control villages, surpassing the learning target set by the Development Impact Bond by 60%. The effects of Educate Girls' program on learning gains were large and statistically significant over the three-year program: Students in EG schools gained on average an additional 1.08 learning levels, or 28%, compared to students in control schools.

Learning gains were higher for treatment students than for control students across all grades and subjects, with relatively higher gains in Math and English than in Hindi and relatively larger treatment effects among students who were exposed to the program for more years. EG's program in Year 3 was particularly effective in increasing test scores.

By the end of the three-year project, Educate Girls had enrolled 768 out-of-school girls, representing 92% of all identified out-of-school school girls eligible for enrollment. Educate Girls thus exceeded the enrollment target of 79% by 16%.

<sup>&</sup>lt;sup>20</sup> Girls are eligible for enrollment if they are between 7 and 14 years old, live in treatment villages, and have not previously been reported enrolled by Educate Girls.

<sup>&</sup>lt;sup>21</sup> In Year 3 EG reported enrolling 155 girls, including four girls whose enrollment in the Rajasthan State Open School (RSOS) will be verified in July 2018. IDinsight was able to verify 148 of the Year 3 enrollments for an error rate of 2%, well below the threshold of 10%. Hence, all 155 girls reported by EG are counted towards the target.



# References

- Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122(3), 1235-1264.
- Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review*, 102(4), 1241-78.
- Muralidharan, K., & Sundararaman, V. (2011). Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, 119(1), 39-77.



# Appendix

#### Appendix 1: Description of Educate Girls' Intervention

#### Enrollment

Educate Girls delivers a comprehensive community intervention to enroll girls into school. This intervention includes identification of out-of-school girls through door-to-door surveys, explanation of the value of schooling to their parents and to the community, and multichannel engagement with households with unenrolled girls. Educate Girls also uses multiple interventions to improve school attendance and prevent drop-outs, such as frequent parent counselling sessions and working with School Management Committees to improve school infrastructure. It also identifies girls who have dropped out and works with the community to re-enroll them into school.

#### Learning

Educate Girls trained volunteers to deliver a child-centric curriculum one to five times a week to boys and girls in Grades 3-5. Volunteers were often drawn from the villages in which they worked. They were incentivized with a small number of skill and career development opportunities, such as free English classes and the possibility of being hired by EG in the future.

In Year 3, EG rolled out a new curriculum called "Gyan Ka Pitara" ("Knowledge Box"). As part of this new curriculum, EG increased the number of teaching sessions per day and conducted home visits to reach students who were frequently absent from school or who needed remedial tutoring. In addition to the thrice yearly rounds of student assessments conducted previously in Years 1 and 2, EG conducted three additional rounds of ASER assessments in Year 3. These additional assessments led EG to identify areas of improvement, which informed adjustments to the clustering of schools for program implementation, the training of volunteers, and the content of remedial classes. School teachers were also more involved in programming in Year 3 through school meetings and block review meetings.



## **Appendix 2: Description of Student Types**

The Evaluation Design Memo outlines five student types, which together make up the full population of students assessed in the evaluation. The interpretation of student types slightly deviates from what is suggested in the Evaluation Design Memo,<sup>22</sup> but was held constant throughout the three Endline data collection exercises and analyses.

Consolidated Student Group	Student Type	Status at Baseline	Status at Endline
	Type I	Enrolled in Grades 1-5	Enrolled, present at school, assessed
Present at	Type II	Enrolled in Grades 1-5	Enrolled, absent at school, assessed
Baseline Type III		Enrolled in Grades 1-5	Not assessed (enrolled or unenrolled, present or absent) <sup>23</sup>
Students	Type IV	Absent or unenrolled	Enrolled, present at school, assessed
Absent at Baseline	Type V	Absent or unenrolled	Enrolled, absent at school, assessed <sup>24</sup>
Newly Enrolled Girls		Unenrolled	Enrolled by EG; present or absent

#### Appendix 3: Data Collection for the Year 3 Endline

IDinsight conducted the third and final Endline between February 2 to February 28, 2018, according to the following protocol:

- IDinsight visited a total of 325<sup>25</sup> schools.<sup>26</sup>
- Out of a sample of 8,237 students (4,211 in treatment, 4,026 in control), we successfully assessed 7,655 students in grades 3-5, or 93% of all sampled students (92% in treatment, 93% in control).
- We also assessed 198 newly enrolled girls in grades 3-5, representing 73% of the newlyenrolled girl population of 272 girls eligible for assessment in Year 3.<sup>27</sup>
- 74% of students were assessed at the school while 26% were assessed at their home. In the majority of cases in which we were not able to assess a child at their home, it was

<sup>&</sup>lt;sup>22</sup> For example, Type III students are considered to be students who drop out from the sample (i.e., their last Endline score is not available) rather than students who dropped out of school.

<sup>&</sup>lt;sup>23</sup> Most students not assessed at Endline are students who dropped out from school and permanently or temporarily migrated. However, students enrolled in school were also sometimes unable to be assessed (for example, if the child was ill or the child or family did not consent to being assessed).

<sup>&</sup>lt;sup>24</sup> Some Type IV/V students may not have been assessed during their last Endline. As with Type III students, we include their latest available score in the calculation of learning gains.

<sup>&</sup>lt;sup>25</sup> There were 332 schools in the original sample. In two cases, treatment and control schools merged. Per the Working Group's decision, IDinsight dropped schools affected by treatment/control merges from the sample (a total of four schools). There were three other in-sample merge cases (treatment school closed and merged with another treatment school or control school closed and merged with control school), which reduced the number of schools to be visited by an additional three schools to 325 schools. In these in-sample merge cases, IDinsight found and surveyed the affected students at home or at their new school. For more information on how school merge cases were dealt with, please refer to Appendix 16.

<sup>&</sup>lt;sup>26</sup> In keeping with the pairwise matching design described in the Baseline Report, students in control villages were in most cases assessed in the same week and by the same surveyors as their treatment equivalents to reduce time and surveyor effects.

<sup>&</sup>lt;sup>27</sup> Many girls enrolled by EG dropped out again and/or permanently migrated, making it harder for surveyors to assess them.



because the family had moved temporarily or permanently to areas too far for surveyors to reach.  $^{\rm 28}$ 

• Children were presented with paper copies of the ASER assessment and their answers were recorded on smartphones via the SurveyCTO electronic data collection interface used in the Baseline and previous Endline assessments. Information about school infrastructure and staffing was collected from the headmaster or head teacher in each school or by direct observation.

	Variable	Average (All)	Std. Dev. (All)	Average Treatment	Average Control	p-Value of Difference
	Grade (1-5)	3.09	1.39	3.1	3.08	0.48
	SC or ST caste (fraction of total)	0.48	0.5	0.48	0.48	0.9
	Age	8.2	1.74	8.19	8.21	0.64
	Female (fraction of total)	0.49	0.5	0.49	0.49	0.69
	Hindi Level (1-6)	2.68	1.69	2.66	2.71	0.5
Children	Math Level (1-5)	2.4	1	2.36	2.43	0.11
Present at Baseline	English Level (1-5)	1.91	1.05	1.89	1.93	0.47
Daseime	English Word Comprehension (fraction answering correctly)	0.18	0.39	0.12	0.24	0.11
	English Sentence Comprehension (fraction answering correctly)	0.41	0.49	0.35	0.44	0.39
	Grade (1-5)	1.95	1	1.98	1.93	0.19
Children Absent At Baseline	SC or ST caste (fraction of total)	0.52	0.5	0.5	0.54	0.3
	Age	7.41	1.52	7.42	7.4	0.76
	Female (fraction of total)	0.51	0.5	0.51	0.52	0.61
NT 1	Grade (1-5)	2.92	1.31	-	-	-
Newly Enrolled Girls	SC or ST caste (fraction of total)	0.51	0.5	-	-	-
GIRIS	Age	9.59	2.13	_	_	_

#### **Appendix 4: Descriptive Student Statistics**

Note: The p-value indicates the likelihood of the difference in means between treatment and control being this large (or larger) by random chance if the difference in means was zero. Age and grade of students absent at Baseline and newly enrolled girls were imputed by subtracting the number of years passed since Baseline. For example, an 8-year-old child in grade 3 during Year 3 Endline is shown as a 6-year-old child in grade 1 in this table.

<sup>&</sup>lt;sup>28</sup> If available, we use the most recent assessment of these children for the calculation of learning gains.



Crada at	Voors of ourocure	Average learning gains					
Baseline	Baseline to EG program		Control students	Difference	Difference (std effects)	p-Value	
1	1	5.97	4.59	1.38	0.46	< 0.01	
2	2	6.76	5.40	1.35	0.41	< 0.01	
3	3	6.13	4.43	1.71	0.50	< 0.01	
4	2	3.59	3.06	0.52	0.16	< 0.01	
5	1	1.32	0.84	0.48	0.28	< 0.01	
	Total	4.96	3.88	1.08	0.31	< 0.01	

#### **Appendix 5: Average Treatment Effects as ASER Levels and Standardized Effects**

Note: Treatment effects are presented as raw differences in scores and as standardized effect sizes. Standardized differences are calculated by subtracting the control mean and dividing by the control standard deviation for each grade. Standardized effects reflect the magnitude of gains in the treatment group relative to the distribution of learning gains and are useful for benchmarking treatment effects against impact estimates from outside programs. The p-value indicates the likelihood of the difference in means between treatment and control being this large (or larger) by random chance if the treatment effect was zero.

#### Appendix 6: Average Treatment Effects by Baseline Grade and Student Type

Grade	All students		All students Present at Baseline		Absent at Baseline	
	Difference	p-Value	Difference	p-Value	Difference	p-Value
1	1.38	< 0.01	1.44	< 0.01	1.36	< 0.01
2	1.35	< 0.01	1.55	< 0.01	1.23	< 0.01
3	1.71	< 0.01	1.70	< 0.01	1.72	< 0.01
4	0.52	< 0.01	0.69	< 0.01	0.39	0.36
5	0.48	< 0.01	0.48	< 0.01	-	-
Total	1.08	< 0.01	1.07	< 0.01	1.26	< 0.01

Note: "Difference" shows the raw difference in learning gains between students in treatment villages and students in control villages (treatment-control). The p-value indicates the likelihood of the difference in means between treatment and control being this large (or larger) by random chance if the treatment effect was zero.

#### **Appendix 7: Average Treatment Effects by Subject and Student Type**

Subject	All Students		Present at Baseline		Absent at Baseline	
	Difference	p-Value	Difference	p-Value	Difference	p-Value
Hindi	0.14	0.03	0.14	< 0.01	0.19	0.17
Math	0.44	< 0.01	0.45	< 0.01	0.49	0.00
English	0.50	< 0.01	0.48	< 0.01	0.58	0.00
Total	1.08	< 0.01	1.07	< 0.01	1.26	0.00

Note: The table shows subject-wise average treatment effects for all students. "Difference" shows the raw difference in learning gains between students in treatment villages and students in control villages (treatment - control). The p-value indicates the likelihood of the difference in means between treatment and control being this large (or larger) by random chance if the treatment effect was zero.



Subject	All Students		Present at Baseline		Absent at Baseline		
	Difference	p-Value	Difference	p-Value	Difference	p-Value	
Hindi	0.20	0.04	0.21	0.02	0.21	0.13	
Math	0.59	< 0.01	0.66	< 0.01	0.53	< 0.01	
English	0.68	< 0.01	0.72	< 0.01	0.65	< 0.01	
Total	1.47	< 0.01	1.59	< 0.01	1.39	< 0.01	

#### **Appendix 8: Average Treatment Effects by Subject and Student Type for Year 3**

Note: The table shows subject-wise average treatment effects for students assessed in the Year 3 Endline (students in Grades 1Y1, 2Y1, and 3Y1). "Difference" shows the raw difference in learning gains between students in treatment villages and students in control villages (treatment - control). The p-value indicates the likelihood of the difference in means between treatment and control being this large (or larger) by random chance if the treatment effect was zero.

#### **Appendix 9: Total Aggregate Learning Gains from Baseline for All Student Types**

	By Year 1 Endline	By Year 2 Endline	By Year 3 Endline
Total	1,461	2,895	8,940
Share of Target (5,592)	26%	52%	160%

Note: Results by Year 1 and Year 2 slightly deviate from the results reported after the Year 2 Endline (2,812 learning by Year 2, 1,498 by Year 1), reflecting updates made in Year 3 as per Appendix 14.

Grade	Year 1 Difference	Year 2 Difference	Year 3 Difference
at Daseinie	Dura un		nom basenne
	Presen	t at Baseline, Types I-III	
1			856
2		162	877
3	237	642	1905
4	400	949	
5	549		
Total	1,186	2302	5136
	Absent	at Baseline, Types IV-V	
1			920
2		-245	583
3	-	64	938
4	-	31	
5	-	96	
Total	-	-54	2583
	Ne	ewly Enrolled Girls	
1		_	227
2		130	254
3	93	178	401
4	81	238	
5	101		
Total	275	647	1221

#### Appendix 10: Aggregate Learning Gains by Baseline Grade, Year, and Type

Note: Scores in bolded text represent the cohort's final score. While the total aggregate gains are consistent, the subaggregate gains of some student types may differ by one learning gain from the numbers reported in Appendix 10 text due to rounding weighted gains at different steps of the calculation. Appendix 10 represents the final result.



		Grade at Baseline					
		1	2	3	4	5	All
	Population	1044	1213	1843	1878	2009	7989
Present at Baseline,	Sampled	1044	1213	1275	1311	1374	6218
Assessed at Endline	Average treatment effect	1.44	1.58	1.74	0.75	0.48	
Type I-II	p-Value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	
	Aggregate gains	856	828	1862	896	549	4991
	Population	89	77	121	83	29	399
	Sampled	89	77	94	64	24	348
	Population: Assessed in Y1 but not Y2 or Y3			52	67		119
	Sampled: Assessed in Y1 but not Y2 or Y3			43	51		94
Present at Baseline.	Average treatment effect (Y1)			-0.14	0.71		
Not Assessed	p-Value (Y1)			0.69	0.23		
at Endline	Aggregate gains (Y1)			-7	53		46
(Type III)	Population: Assessed in Y2 but not Y3		41	56			97
	Sampled: Assessed in Y2 but not Y3		41	43			84
	Average treatment effect (Y2)		1.54	1.64			
	p-Value (Y2)		0.00	0.06			
	Aggregate gains (Y2)		49	49			98
	Population	1872	1484	838	685	-	4879
Absent at Baseline,	Sampled	1872	1484	569	455	-	4380
Assessed at Endline	Average treatment effect	1.36	1.22	1.72	0.39	-	
(Type IV/V)	p-Value	< 0.01	< 0.01	< 0.01	0.36	-	
	Aggregate gains	920	651	969	31	96	2667
	Population	63	89	153	113		418
	Sampled	63	89	96	74		322
Absent at Baseline,	Population: Assessed in Y2 but not Y3		59	63			
at Endline	Sampled: Assessed in Y2 but not Y3		59	37			
(Type IV/V)	Average treatment effect (Y2)		-0.34	0.57			
	p-Value		0.91	0.10			
	Aggregate gains (Y3)		-68	-15			-83
	Population/Sampled	71	103	98	88	61	421
	Assessed in Y1 but not Y2 or Y3			1	2	52	55
	Average treatment effect (Y1)			1.00	6.00	1.94	
	Aggregate Gains (Y1)			1	12	101	114
Newly Enrolled	Assessed in Y2 but not Y3		15	8	75	0	98
Girls	Average treatment effect (Y2)		1.20	1.63	3.01		
	Aggregate Gains (Y2)		18	13	226		257
	Assessed in Y3	48	72	78			197
	Average treatment effect (Y3)	4.73	3.28	4.96			
	Aggregate Gains (Y3)	227	236	387			850
TOTAL	Aggregate Gains						8940

#### Appendix 11: Breakdown of Learning Gains from Baseline by Grade and Type

Note: The calculated learning gains in this table represent the final result. The sub-aggregate gains of some student types may differ by one learning gain in other tables due to rounding weighted gains at different steps of the calculation of aggregate learning gains. In Year 1, the Working Group decided to impute learning gains for students not present at Baseline in grade 5 since they were not included in the Year 1 sample and would have graduated from the program in Year 2. The Working Group agreed to err on the side of overestimating learning gains for this group by assuming that the effect of Educate Girls' program on students not assessed at Baseline in grade 5 were the same as the effect on students assessed at Baseline.



	Subgroup	Average (Treatment)	Average (Control)	Difference	p-Value of Difference
	General	5.90	4.58	1.31	< 0.01
	OBC	4.78	3.86	0.91	< 0.01
Caste Category	SC	5.14	3.88	1.26	< 0.01
	ST	4.64	3.78	0.86	< 0.01
	Boy	5.02	3.98	1.04	< 0.01
Gender	Girl	4.90	3.77	1.13	< 0.01
	Bijoliya	4.98	3.35	1.62	< 0.01
Block	Jahajpur	5.00	4.22	0.78	< 0.01
	Mandalgarh	4.91	3.77	1.14	< 0.01
	1	2.01	1.86	0.15	0.09
	2	2.07	1.91	0.16	< 0.01
Hindi Score	3	1.86	1.50	0.36	0.06
at Baseline	4	1.17	1.06	0.10	0.08
	5	0.61	0.60	0.01	0.73
	6	-0.25	-0.16	-0.10	0.03
	1	1.82	1.36	0.46	< 0.01
	2	1.39	0.91	0.49	< 0.01
Math Score	3	1.04	0.62	0.42	< 0.01
at Dasenne	4	0.42	0.26	0.16	0.02
	5	-0.13	-0.33	0.20	< 0.01
	1	1.76	1.19	0.57	< 0.01
-	2	1.40	0.97	0.42	< 0.01
English Score	3	0.83	0.48	0.35	< 0.01
at Daseinie	4	0.50	0.13	0.37	0.03
	5	-0.23	-0.30	0.07	0.63
Total		4.01	2.94	1.07	< 0.01

#### Appendix 12: Sub-Group Analysis by Caste, Grade, Block, and Baseline Scores

Note: Newly-enrolled girls are omitted from all analyses and students absent at baseline are omitted from the analysis of performance at baseline. For the subgroup analyses by caste category, gender, and block, mean values represent total learning gains (across all subjects). As a reminder, students who are absent at baseline are imputed the lowest possible score (3 out of 16 points), which explains the high learning gains for subgroups including those students (since this imputation is done for both Treatment and Control students it does not affect the unbiasedness of the ATE estimator). For the subgroup analyses by baseline scores, mean values represent learning gains in the respective subject (for students present at baseline). The p-values in this table are the likelihoods that, if the treatment effect is zero, then the difference in means between treatment and control could be this large by random chance.

#### **Appendix 13: Assessment Location of Students**

Student Type	At School	At Home
Students Present at Baseline	78%	22%
Students Absent at Baseline	73%	27%



	By Year 1 Endline	By Year 2 Endline	By Year 3 Endline
A: Girls Enrolled	322	613	768
B: Girls Eligible for Enrollment	744	835	837
C: Share of Girls Enrolled against Final Target (837 Girls)	38%	73%	92%
D: Share of Target (D=C/79%)	48%	92%	116%

#### Appendix 14: Newly Enrolled Girls since Baseline

#### Appendix 15: Changes to Year 1 and Year 2 Results

IDinsight made updates to the data from Year 1 and Year 2, leading to small changes in the calculated aggregate learning gains by Year 1 and 2. These changes represent 0.01% (Year 1) and 2.1% (Year 2) of the final target.

- Students in grades 4 and 5 at Baseline were expected to progress to grades 6 and 7 by Year 3. However, 32 students from Baseline grades 4 and 5 were still in grades 3-5 at the time of the Year 3 Endline, and thus assessed this year. Likewise, two students from Baseline grade 5 were still in grade 5 during the Year 2 Endline. We included these assessments in the final calculation of learning gains, leading to changes in the learning gains of students in grades 4<sub>Y1</sub> and 5<sub>Y1</sub> despite these cohorts generally not being part of the Year 3 student assessments.
- 26 children present at Baseline subsequently dropped out of school and were later enrolled by EG. We shifted these students from Type I-III to the Newly Enrolled Girls category. Since 100% of Newly Enrolled Girls were sampled, their sampling weight was changed to 1. The remaining Type I-III students in the cohorts from which these students were removed kept their original sampling weights.
- During the third round of student assessments, we identified 64 students who were listed twice on our student lists. While none of them have been assessed twice, removing these duplicates affects sampling weights.
- We made updates to school assignments for several students who were incorrectly attributed to schools with similar names (e.g., Ragunathpura vs. Ragunathpra and Rampuriya vs. Rampuria) leading to small changes in sampling weights.



Variable (* indicates average if answer to preceding question is "ves")	Average (All)	Std. Dev. (All)	Average (Treatment)	Average (Control)	p-Value of Difference
# of Headmasters (Appointed)	0.32	0.47	0.31	0.32	0.97
# of Headmasters (Observed)	0.3	0.46	0.31	0.29	0.69
# of Teachers (Appointed)	2.88	1.82	3.01	2.75	0.24
# of Teachers (Observed)	2.52	1.76	2.6	2.44	0.53
# of Parateachers (Appointed)	0.12	0.41	0.11	0.12	0.79
# of Parateachers (Observed)	0.08	0.28	0.08	0.08	0.85
Existence of SMC	1	0	1	1	
# of SMC Members*	11.34	3.68	11.5	11.18	0.53
# of SMC Meetings*	12.6	55.39	15.55	9.54	0.32
Mid-Day Meal Served	0.84	0.37	0.87	0.82	0.23
School Kitchen Available	0.95	0.21	0.95	0.96	0.78
Observed Food Served	0.84	0.37	0.83	0.84	0.74
Evidence of Mid-Day Meal	0.73	0.44	0.73	0.73	0.97
# of Pucca (Permanent) Rooms	5.69	2.47	5.64	5.74	0.69
# of Rooms for Teaching	2.56	1.83	2.55	2.57	0.94
Play Area	0.72	0.45	0.7	0.74	0.43
Usable Equipment in Play Area	0.15	0.36	0.15	0.15	0.86
Sports Equipment	0.91	0.89	0.8	1.03	0.02
Library Books	0.75	0.83	0.79	0.71	0.39
Children Using Books*	0.7	0.46	0.71	0.68	0.77
Handpump/Tap	0.66	0.47	0.69	0.64	0.25
Handpump in Usable Condition*	0.83	0.37	0.84	0.83	0.93
Source of Drinking Water	0.46	0.57	0.43	0.48	0.4
Electricity	0.46	0.5	0.44	0.47	0.61
Electricity at Visit*	0.66	0.47	0.67	0.66	0.95
School Wall/Boundary	0.5	0.5	0.5	0.5	0.96
Computers	0.03	0.18	0.04	0.03	0.99
Children Using Computers*	0.27	0.47	0.17	0.4	0.14
Tables and Chairs Available	0.06	0.24	0.06	0.06	0.8
Dari (Carpet) for Seating	0.97	0.16	0.97	0.97	0.98
Usable Blackboard	1	0	1	1	
Other Learning Materials in Classroom	0.92	0.27	0.93	0.9	0.42
Common Toilet	0.21	0.41	0.18	0.24	0.13
Girls Toilet	0.94	0.23	0.95	0.94	0.64
Boys Toilet	0.93	0.26	0.92	0.93	0.66
Total Enrollment Grades 1 to 5	45.18	23.31	45.63	44.71	0.79

# Appendix 16: Descriptive Statistics of Schools Surveyed in Year 3

Note: Data from 320 schools from Year 3 Endline. The p-values in this table are the likelihoods to observe differences in means between treatment and control this large (or larger) by random chance if there were no mean differences between treatment and control schools.



## **Appendix 17: Merged Schools**

#### **Treatment-Control Merge Cases**

As per the Working Group's decision from 2017, in cases where a treatment school closed and merged with a control school or a control school closed and merged with a treatment school, students have not been assessed after the school merge occurred. However, all learning gains that were captured before the schools merged are included in the calculation of outcome payments.

i reatment school closed and merged with control school					
School	DISE Code	School Merged With	DISE Code	Year	
G.P.S. NAYA GAU	8241011209	G.P.S.JORA JI KA KHERA	8241011601	Y2	
Control school closed and merged with treatment school					
School	DISE Code	School Merged With	DISE Code	Year	
G.P.S. HIMMAT PURA	8241115802	G.P.S. BHEROO KA RADHA	8241115902	Y2	

## school closed and merged with control school

#### **Out-of-Sample Merge Cases**

In cases where an in-sample school merged with an out-of-sample school, we continued to assess all sampled students from the in-sample school. IDinsight did not assess any students that were previously enrolled in out-of-sample schools.

School	DISE Code	School Merged With	Year	
G.P.S. GOPALPURA	8241000106	G.P.S. MAGANPURA	Y2	
G.P.S. MANAK CHOUK	8241107804	G.G.U.P.S. MAHUO	Y2	
G.P.S NANA BABA KA JHUPRA	8241009202	G.P.S. BHAIRU KA KHERA	Y2	
G.P.S. PIPALDA	8241108802	G.S.K.P.S. RAMPURIYA	Y3	
Control school c	losed and merged w	ith out-of-sample school		
School	DISE Code	School Merged With	Year	
G.P.S. KANJORA KA JOPARA	8241102603	G.S.S.S. RAJGARH SARTHALA	Y3	
G.P.S. BAGTHALA	8241100801	G.S.S.S. RAJGARH	Y3	
Out-of-sample school closed and merged with in-sample school				
School	School Merged With	DISE CODE	Year	
G.P.S. LAXMIPURA	G.U.P.S. DAGARIYA	8241028401	¥3	
G.P.S. RATANPURA	G.P.S. JAJARPURA	8241116801	Y3	

#### Treatment school closed and merged with out-of-sample school

#### **In-Sample Merge Cases**

In cases where a treatment school merged with another treatment school or a control school merged with another control school, IDinsight continued to assess all sampled students from both schools.

Treatment school closed and merged with another treatment school					
School	DISE Code	School Merged With	DISE Code	Year	
G.P.S. BHIL BASTI	8241112602	G.P.S. MEENA KA JHONPARIYA	8241111606	Y2	
G.P.S. BILIYA KA JOPHDA	8241102002	G.U.P.S. BILIYA	8241102001	Y2	
Control sch	ool closed and	merged with another cont	rol school		
School	DISE Code	School Merged With	DISE Code	Year	
G.P.S. BHARJI KA KHERA	8241104601	G.U.P.S. SHAKTA JI KA KHERA	8241104801	Y2	



## Appendix 18a: ASER Testing Tool for Hindi in Year 3 Endline



## Appendix 18b: ASER Testing Tool for Math in Year 3 Endline

Number recognition	
1 – 9	

SSMENT (Version A): LEVELS 0-4

1	4
7	3
6	9
5	2

nents except for Hindi Level 5 developed by the ASER Centre (www.asercentre.org)

Number recognition 10 – 99				
52	83			
37	27			
55	28			
91	65			
36	43			



All assessments except for Hindi Level 5 developed by the ASER Centre (www.asercentre.org)



## Appendix 18c: ASER Testing Tool for English in Year 3 Endline





All assessments except of Hindi Level 5 developed by ASER Centre (www.asercentre.org)

ENGLISH ASSESSMENT: LEVELS 0-4

## Appendix 19: Map of Bhilwara District



Note: "Beejoliya" and "Jahazpur" are alternative spellings of Bijoliya and Jahajpur, respectively.