Supplementary Online Appendix

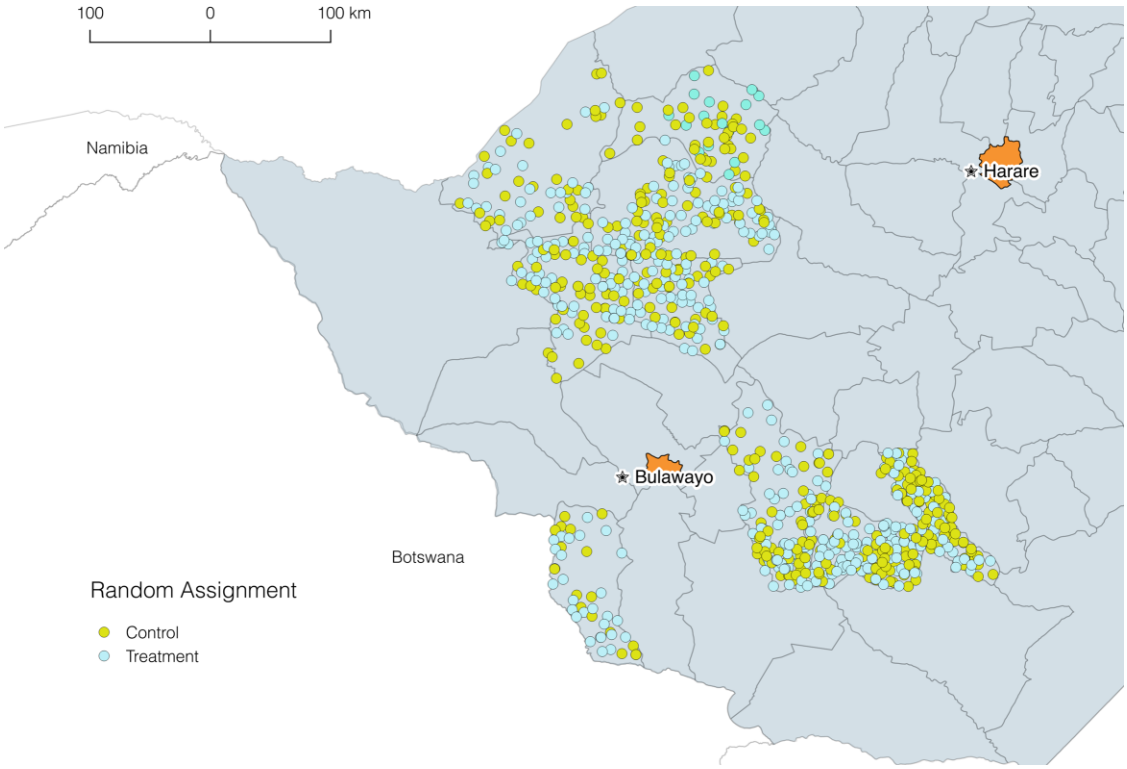
Can Discussions about Girls' Education Improve Academic Outcomes? Evidence from a Randomized Development Project

Christopher S. Cotton, Ardyn Nordstrom, Jordan Nanowski, and Eric Richert

S1. School Locations

Figure S1.1 shows a map of the Improving Girls' Access through Transformative Education (IGATE) school locations across rural districts in Zimbabwe.

Figure S1.1. IGATE School Locations



Source: Generated by the authors.
Note: The eligible schools in intervention districts were randomly assigned to be in the treatment and control groups. This figure shows the location of these schools across Zimbabwe.

S2. EGRA/EGMA Test Details

As is standard with the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) assessments, the questions in each test were described verbally, one by one, by a professional enumerator to individual students. Students then provided their answers verbally and enumerators record whether the student's answer was correct. During the test, students are given visual stimuli to follow along and to see the specific numbers, letters, and words they are asked to say or analyze. There are five subtasks that make up the numeracy assessment: number identification, number quantities, missing numbers, addition, and subtraction. The number identification subtask consists of 20 numbers which students are asked to identify in one minute. An example of a typical EGMA number identification subtask as viewed by the enumerator is shown in fig. S2.1.

In the quantity discrimination subtask, a student is presented with a list of 10 pairs of numbers and is asked to identify the larger number. This exercise is not timed but ends after four incorrect answers in a row or a hesitation of 5 seconds by the student. This stop rule trigger is shown in fig. S2.2.

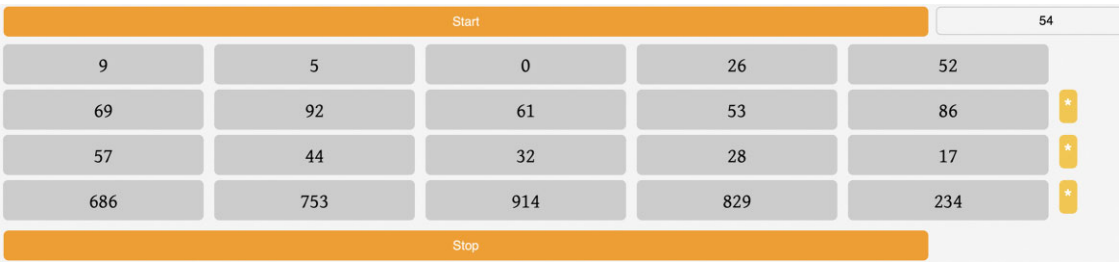
The addition and subtraction level 1 components include 20 problems each. According to the EGMA guidelines, subtraction questions must be the inverse of the additional questions. A stop rule after five incorrect answers applies to these tasks as well.

There are five literacy subtasks: letter and sound identification, invented words, oral fluency (grades 1–5, and grades 6–7), and reading comprehension. The letter and sound identification tasks involve students phonetically reading individual letters in the alphabet, much like the number identification task. The enumerator records each correct pronunciation. Invented-word tasks involve 50 words that do not have a meaning in English or in the local languages. The student is asked to read each made-up word aloud and the enumerator records each correct pronunciation. An example of an invented words subtask as viewed by the enumerator is shown in fig. S2.3.

The remaining subtasks, oral fluency and reading comprehension, ask students to read a short story aloud. Enumerators are instructed to record the words the students misidentified or mispronounced and to identify the last word the student correctly said aloud within the time limit. The reading comprehension task then asks the students questions about the passage to assess their understanding of the story they just read.

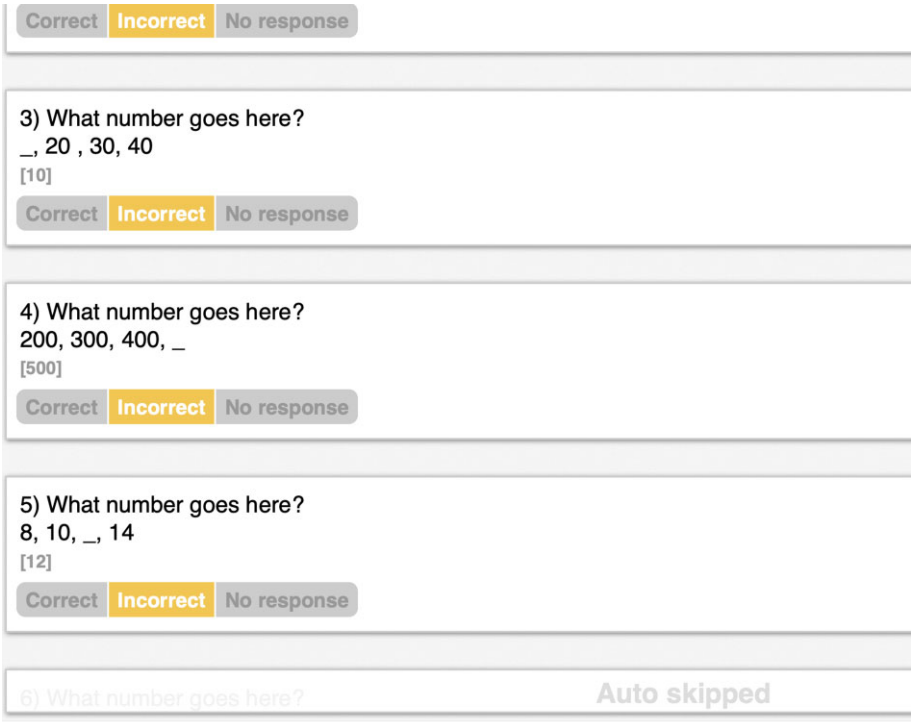
The test design guidelines specify all details about each question's difficulty level. This includes details about the number each sequence increases by in numeracy subtasks and the number of single-, double-, and triple-digit numbers to be used in the missing numbers and number identification subtasks; subtraction problems are required to be the inverse of the addition problems. In the first two EGRA components, the versions are differentiated by reordering letters or words within the rows to retain the same level of difficulty. The EGRA story subtasks are written with the intention of maintaining the same difficulty using

Figure S2.1. Early Grade Mathematics Assessment (EGMA) 1 (Number Identification) Example



Source: Screenshot of the instrument in the Tangerine survey platform, taken by the authors.
Note: The number identification subtask shown in this figure allows enumerators to record which numbers the student could identify in one minute. Students were given a physical copy of this table, and enumerators selected the identified numbers.

Figure S2.2. Early Grade Mathematics Assessment (EGMA) 3 (Missing Numbers) Example: Early Stop Rule Trigger



Source: Screenshot of the instrument in the Tangerine survey platform, taken by the authors.
Note: In the survey tool, if enumerators recorded four incorrect answers in a row, the remaining questions in this subtask are skipped. This stop rule trigger is shown here.

Figure S2.3. Early Grade Reading Assessment (EGRA) 2 (Invented Words) Example

Source: Screenshot of the instrument in the Tangerine survey platform, taken by the authors.
Note: As part of the invented word subtask in the Early Grade Reading Assessment (EGRA), students are asked to pronounce a set of made up words. Students are given a physical version of the table shown here, and enumerators mark which words students correctly pronounced.

the same number of words per sentence and per passage and using a similar vocabulary. Given this strict structure, different versions of the tests are not likely to have different difficulties.

S3. Sample Attrition

This section examines the attrition rates to assess the balance between the treatment and control groups over time.

To confirm that attrition is not systematically related to treatment status, treatment status is used to predict attrition. The results of this regression are presented in table S3.1 and suggest that treatment status does not predict attrition status at either baseline or midline.

It is possible that caregivers with no formal education would be inclined to migrate out of treated areas, which could lead to systematic differences in attrition. However, table S3.2 shows that having an uneducated caregiver does not significantly affect a girl's likelihood of attriting.

Tables S3.3 and S3.4 also show that the baseline characteristics and test scores are similar between individuals who could be recontacted, and those who attrited by midline. Together, these findings suggest that attrition is not related to treatment status in any way that would impact our identification strategy.

However, as a test, fig. S3.1 shows the results of a bounding exercise using Lee bounds (Tauchmann 2014). While the upper bound for the impact on mathematics score at midline does significantly deviate from zero, the lower bound does not significantly deviate from zero. Like other bounding techniques, it can be difficult to interpret the findings of Lee trimming exercises (Glennerster and Takavarasha 2014). This may suggest that attrition plays a role in the impact observed in mathematics. However, it should

Table S3.1. Attrition Balance

	Attrition at midline	Attrition at endline
Treatment	0.0285	0.0406
	(0.0213)	(0.0255)
Controls	✓	✓
Observations	984	984
R-squared	0.040	0.016

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: The table reports the coefficient on an indicator for belonging to a treatment school. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S3.2. Attrition and Caregiver Education

	Attrition at midline		Attrition at endline	
	(1)	(2)	(3)	(1)
Caregiver has no education	0.0205	0.0396	0.0120	−0.0413
	(0.0281)	(0.0371)	(0.0342)	(0.0610)
Treatment	–	0.0314	–	0.0322
		(0.0222)		(0.0280)
Treatment * caregiver has no education	–	−0.0308	–	0.0848
		(0.0542)		(0.0736)
Controls		✓		✓
Observations	975	975	975	975
R-squared	0.001	0.041	0.000	0.017

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: The table reports the coefficient on an indicator for belonging to a treatment school. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S3.3. Summary Statistics by Attrition Status

	Attrited by midline			Attrited by endline			Recontacted at midline		
	Control	Treatment	Difference	Control	Treatment	Difference	Control	Treatment	Difference
Age	9.75	8.892	−0.858	9.592	9.552	−0.04	9.377	9.38	0.003
	(2.261)	(1.499)		(2.179)	(1.913)		(2.011)	(1.997)	
Grade	3.583	3.143	−0.44	3.918	3.729	−0.189	3.676	3.627	−0.049
	(1.730)	(1.353)		(1.766)	(1.579)		(1.746)	(1.779)	
Illness	0.091	0.143	0.052	0.082	0.094	0.012	0.103	0.112	0.009
	(0.302)	(0.356)		(0.277)	(0.293)		(0.305)	(0.316)	
Disability	0.167	0.286	0.119	0.143	0.177	0.034	0.174	0.191	0.017
	(0.389)	(0.460)		(0.354)	(0.384)		(0.380)	(0.394)	
Orphan	0.143	0.000	−0.143	0.106	0.014	−0.092	0.0676	0.0536	−0.014
	(0.378)	(0.000)		(0.312)	(0.116)		(0.252)	(0.226)	
Travel time to school (minutes)	28.500	37.933	9.433	27.486	31.313	3.827	32.96	35.12	2.16
	(13.754)	(31.800)		(14.963)	(24.300)		(23.160)	(27.300)	
Caregiver has no education	0.167	0.143	−0.024	0.061	0.137	0.076	0.0676	0.0979	0.0303
	(0.389)	(0.356)		(0.242)	(0.346)		(0.252)	(0.298)	
Caregiver has primary education	0.583	0.464	−0.119	0.388	0.453	0.065	0.523	0.55	0.027
	(0.515)	(0.508)		(0.492)	(0.500)		(0.500)	(0.498)	
Caregiver has secondary education	0.250	0.399	0.149	0.551	0.411	−0.14	0.409	0.352	−0.057
	(0.452)	(0.497)		(0.503)	(0.494)		(0.493)	(0.478)	
Caregiver works outside the household	0.167	0.250	0.083	0.306	0.219	−0.087	0.221	0.235	0.014
	(0.389)	(0.441)		(0.466)	(0.416)		(0.415)	(0.425)	
Household has a TV	0.250	0.143	−0.107	0.184	0.0729	−0.111	0.164	0.152	−0.012
	(0.452)	(0.356)		(0.391)	(0.261)		(0.371)	(0.359)	
Observations	12	28	16	49	96	47	385	557	172

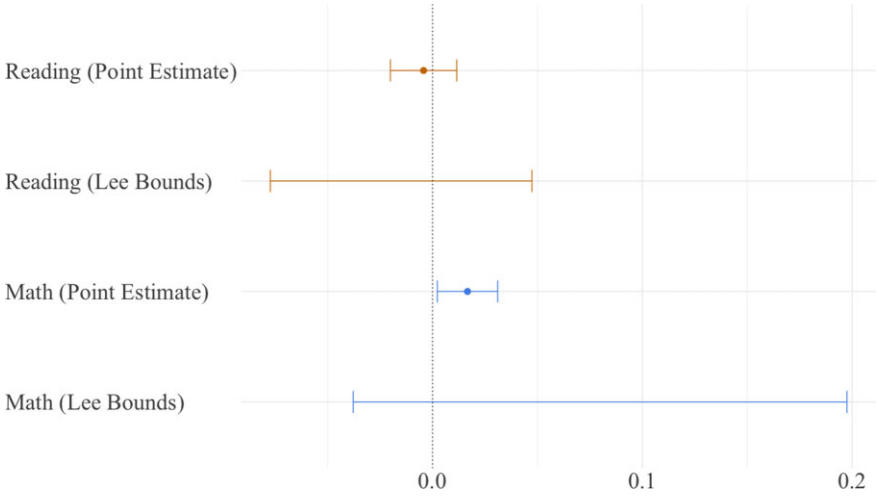
Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: Note these numbers measure baseline levels for girls with different attrition statuses at midline and endline.

Table S3.4. Baseline Numeracy Test Subtasks by Attrition Status

	Attrited by midline			Attrited by endline			Recontacted at midline		
	Control	Treatment	Difference	Control	Treatment	Difference	Control	Treatment	Difference
EGMA 1: Number identification	0.671	0.607	−0.064	0.735	0.718	−0.017	0.693	0.677	−0.016
	(0.324)	(0.330)		(0.328)	(0.325)		(0.327)	(0.338)	
% of questions attempted	0.963	0.928	−0.035	0.929	0.958	0.029	0.951	0.955	0.004
	(0.130)	(0.117)		(0.174)	(0.127)		(0.124)	(0.122)	
% of time spent	0.776	0.868	0.092	0.699	0.745	0.046	0.733	0.737	0.004
	(0.202)	(0.184)		(0.240)	(0.230)		(0.236)	(0.236)	
EGMA 2: Number quantities	0.492	0.430	−0.062	0.643	0.615	−0.028	0.571	0.557	−0.014
	(0.284)	(0.324)		(0.346)	(0.341)		(0.331)	(0.354)	
% of questions attempted	0.875	0.844	−0.031	0.923	0.908	−0.015	0.902	0.876	−0.026
	(0.218)	(0.233)		(0.176)	(0.201)		(0.195)	(0.224)	
EGMA 3: Missing numbers	0.367	0.326	−0.041	0.428	0.390	−0.038	0.376	0.372	−0.004
	(0.206)	(0.233)		(0.281)	(0.245)		(0.260)	(0.266)	
% of questions attempted	0.867	0.815	−0.052	0.872	0.867	−0.005	0.832	0.820	−0.012
	(0.227)	(0.228)		(0.209)	(0.196)		(0.222)	(0.232)	
EGMA 4: Addition	0.358	0.296	−0.062	0.465	0.429	−0.036	0.427	0.409	−0.018
	(0.231)	(0.237)		(0.261)	(0.261)		(0.294)	(0.284)	
% of questions attempted	0.729	0.624	−0.105	0.700	0.665	−0.035	0.710	0.688	−0.022
	(0.129)	(0.0892)		(0.134)	(0.144)		(0.152)	(0.161)	
% of time spent	1.000	1.000	0.000	0.990	0.996	0.006	0.986	0.984	−0.002
	(0.000)	(0.000)		(0.0520)	(0.0298)		(0.0630)	(0.0776)	
EGMA 5: Subtraction	0.225	0.228	0.003	0.374	0.294	−0.08	0.316	0.298	−0.018
	(0.229)	(0.199)		(0.263)	(0.231)		(0.273)	(0.250)	
% of questions attempted	0.721	0.602	−0.119	0.684	0.635	−0.049	0.668	0.649	−0.019
	(0.147)	(0.110)		(0.149)	(0.155)		(0.146)	(0.161)	
% of time spent	1.000	0.985	−0.015	0.986	0.996	0.010	0.986	0.990	0.004
	(0.000)	(0.0802)		(0.0726)	(0.0315)		(0.0683)	(0.0629)	
Average numeracy score	0.423	0.377	−0.046	0.529	0.489	−0.04	0.477	0.462	−0.015
	(0.220)	(0.236)		(0.264)	(0.243)		(0.265)	(0.266)	
% of questions attempted	0.831	0.763	−0.068	0.822	0.806	−0.016	0.813	0.798	−0.015
	(0.0738)	(0.0968)		(0.0946)	(0.101)		(0.111)	(0.114)	
Observations	12	28	16	49	96	47	385	557	172

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: Note these numbers measure baseline levels for girls with different attrition statuses at midline and endline. EGMA refers to Early Grade Mathematics Assessment.

Figure S3.1. Treatment Effects (Baseline to Midline)—Bounded vs Unbounded



Source: Generated by the authors.
Note: This figure shows the results of a bounding exercise using Lee bounds. While the upper bound for the impact on mathematics score at midline does significantly deviate from zero, the lower bound does not significantly deviate from zero.

be acknowledged that at midline, the attrition rate in the treatment and control groups is not statistically significant, which limits the usefulness of this approach.

S4. Additional Treatment and Control Comparison Tables

This appendix compares the treatment and control group baseline characteristics, illustrating that there are no substantial differences between the groups ahead of the IGATE implementation.

Table S4.1. Baseline Numeracy Test Score by Grade

Baseline grade	Control	Treatment
Grade 1	0.143	0.111
	(0.173)	(0.149)
Grade 2	0.324	0.321
	(0.178)	(0.176)
Grade 3	0.407	0.466
	(0.208)	(0.212)
Grade 4	0.589	0.556
	(0.178)	(0.205)
Grade 5	0.638	0.626
	(0.181)	(0.207)
Grade 6	0.66	0.626
	(0.208)	(0.214)
Grade 7	0.769	0.674
	(0.167)	(0.205)

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: These numbers measure baseline levels for girls who could be recontacted by midline.

Table S4.2. Numeracy Test Subtasks—Summary Statistics

	Control	Treatment	Difference
EGMA 1: Number identification	0.693	0.677	−0.016
	(0.327)	(0.338)	
% of questions attempted	0.951	0.955	0.004
	(0.124)	(0.122)	
% of time spent	0.733	0.737	0.004
	(0.236)	(0.236)	
EGMA 2: Number quantities	0.571	0.557	−0.014
	(0.331)	(0.354)	
% of questions attempted	0.902	0.876	−0.026
	(0.195)	(0.224)	
EGMA 3: Missing numbers	0.376	0.372	−0.004
	(0.260)	(0.266)	
% of questions attempted	0.832	0.820	−0.012
	(0.222)	(0.232)	
EGMA 4: Addition	0.427	0.409	−0.018
	(0.294)	(0.284)	
% of questions attempted	0.710	0.688	−0.022
	(0.152)	(0.161)	
% of time spent	0.986	0.984	−0.002
	(0.0630)	(0.0776)	
EGMA 5: Subtraction	0.316	0.298	−0.018
	(0.273)	(0.250)	
% of questions attempted	0.668	0.649	−0.019
	(0.146)	(0.161)	
% of time spent	0.986	0.990	0.004
	(0.0683)	(0.0629)	
Average numeracy score	0.477	0.462	−0.015
	(0.265)	(0.266)	
% of questions attempted	0.813	0.798	−0.015
	(0.111)	(0.114)	
Observations	385	557	

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: Note these numbers measure baseline levels for girls who could be recontacted at midline. EGMA refers to Early Grade Mathematics Assessment.

Table S4.1 shows that the average test score increases as students get older. The Early Grade Mathematics Assessment used in this study is designed to be able to detect improvements within this group, but does not perform as well for students in upper secondary school. This may also explain why improvements in math cannot be detected later in the program, since there is less room for variation.

Tables S4.2 and S4.3 show that there are no significant differences in students' baseline test scores or the distribution of what grade students are in at baseline, respectively.

S4.1. Grade Progression Outcomes

Table S4.4 considers whether the IGATE program has had any impact on progression outcomes. There is no evidence to suggest that IGATE has had an impact on the likelihood that students progress to the next grade. Note that Zimbabwe has an automatic progression policy; however, many students will still repeat a grade during their time in school. This is typically due to low attendance.

Table S4.5 presents the findings from the midline math-test-score analysis, restricting the sample to only girls who are still enrolled in school at midline. Restricting to just the girls who are still enrolled,

Table S4.3. Sample Grade Distribution (Percentage)

Baseline grade	Control	Treatment	Difference
1	15%	15%	0%
2	17%	17%	− 1%
3	14%	18%	3%
4	19%	16%	− 3%
5	21%	20%	− 1%
6	8%	9%	0%
7	5%	6%	0%
Observations	385	557	

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: Note these numbers measure baseline levels for girls who could be recontacted at midline.

Table S4.4. Impact on Grade Progression

	Progression at midline (1)	Progression at endline (2)
Treatment	0.0324	−0.00889
	(0.0272)	(0.0286)
Controls	✓	✓
Observations	942	796
R-squared	0.131	0.108

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: The table reports the coefficient on an indicator for belonging to a treatment school. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table S4.5. Impact on Mathematics Outcomes (Girls Enrolled at Midline)

	Number identification EGMA 1	Number quantities EGMA 2	Missing numbers EGMA 3	Addition EGMA 4	Subtraction EGMA 5	Average Total
Treatment	0.00537	0.0216	0.0123	0.0215	0.00704	0.0147
	(0.0132)	(0.0163)	(0.0131)	(0.0157)	(0.0155)	(0.0111)
Controls	✓	✓	✓	✓	✓	✓
Observations	913	913	913	913	913	913
R-squared	0.325	0.354	0.346	0.427	0.423	0.541

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.
Note: EGMA refers to Early Grade Mathematics Assessment. The table reports the coefficient on an indicator for belonging to a treatment school. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

we find that the program's impact on mathematics is still positive but no longer statistically significant, suggesting that the program's impact on mathematics is not just a consequence of the program's impact on enrollment.

S4.2. Additional Analyses of Impact on Mathematics

Table S4.6 reports results from an analysis considering the impact of IGATE on the number of questions attempted on the EGMA exam. The first panel shows that the relevant coefficients from the DiD analysis of

Table S4.6. Percentage of Questions Attempted

	Number identification EGMA 1	Number quantities EGMA 2	Missing numbers EGMA 3	Addition EGMA 4	Subtraction EGMA 5	Average total
Baseline to midline						
Treatment	0.003	0.010	0.007	0.008	−0.001	0.008
	(0.003)	(0.008)	(0.010)	(0.012)	(0.014)	(0.007)
Controls	✓	✓	✓	✓	✓	✓
Observations	942	942	942	942	942	942
Adjusted *R*-squared	0.040	0.192	0.238	0.324	0.241	0.377
Baseline to endline						
Treatment	0.001	0.003	0.018**	0.018**	−0.0004	0.010
	(0.002)	(0.006)	(0.009)	(0.009)	(0.016)	(0.007)
Controls	✓	✓	✓	✓	✓	✓
Midline attempt percent						
Observations	794	794	794	794	794	794
Adjusted *R*-squared	0.020	0.098	0.158	0.158	0.182	0.288
Baseline to endline						
Treatment	0.001	−0.0002	0.015*	0.003	−0.003	0.003
	(0.002)	(0.005)	(0.008)	(0.011)	(0.014)	(0.005)
Controls	✓	✓	✓	✓	✓	✓
Midline attempt percent	✓	✓	✓	✓	✓	✓
Observations	794	794	794	794	794	794
Adjusted *R*-squared	0.033	0.384	0.353	0.379	0.342	0.595
Timed	Yes	No	No	Yes	Yes	–
Early stop rule	No	Yes	Yes	No	No	–

Source: Authors' analysis based on data from learner surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.

Note: EGMA refers to Early Grade Mathematics Assessment. The table reports the coefficient on an indicator for belonging to a treatment school. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

question attempts are positive and significant. This suggests that the community information interventions led girls to answer more questions between baseline and midline. Note that in the case of the number quantities subtask, this should be interpreted as an indication of improved ability rather than increased effort, since this subtask was stopped after participants incorrectly answered four questions in a row.

S5. Attitude Measures

Table S5.1 reports the proportion of caregivers who report different supportive attitudes at baseline, midline, and endline. No significant differences exist between the treatment and comparison groups.

Table S5.1. Caregiver and Learner Attitudes at Baseline, Midline, and Endline

	Control	Treatment	Difference
Baseline			
Caregiver has positive aspirations for their child	0.979	0.991	0.012
	(0.143)	(0.0945)	
Caregiver believes girls can learn as much as boys	0.979	0.964	−0.015
	(0.143)	(0.186)	
Caregiver listens to daughter's views when making decisions about girl's education	0.662	0.646	−0.016
	(0.474)	(0.479)	
Midline			
Caregiver has positive aspirations for their child	0.966	0.995	0.029
	(0.181)	(0.0684)	
Caregiver believes girls can learn as much as boys	0.973	0.953	−0.02
	(0.162)	(0.212)	
Caregiver listens to daughter's views when making decisions about girl's education	0.591	0.724	0.133
	(0.493)	(0.448)	
Endline			
Caregiver has positive aspirations for their child	0.988	0.998	0.01
	(0.110)	(0.0462)	
Caregiver believes girls can learn as much as boys	0.976	0.968	−0.008
	(0.154)	(0.176)	
Caregiver listens to daughter's views when making decisions about girl's education	0.683	0.642	−0.041
	(0.466)	(0.480)	

Source: Authors' analysis based on data from caregiver surveys collected for the original evaluation of the Improving Girls' Access through Transformative Education (IGATE) program.

Note: The table reports the proportion of caregivers who report different supportive attitudes at baseline, midline, and endline. No significant differences exist between the treatment and comparison groups.